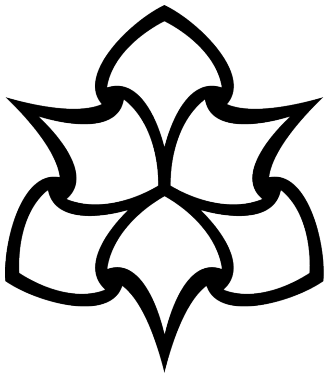


# Production of more stable induced pluripotent stem cells using the Doggybone (dbDNA) vector.

C D THORNTON

PhD 2020

# Production of more stable induced pluripotent stem cells using the Doggybone (dbDNA) vector.

Christopher David Thornton

A thesis submitted in partial fulfilment of the requirements of Manchester Metropolitan University for the degree of Doctor of Philosophy

Department of Life Sciences,

Manchester Metropolitan University

2020

**Abstract:**

The application of induced pluripotent stem cell (iPSC)-derived cells in clinical trials is in its infancy but the potential is vast. A key asset of iPSCs is the ability to derive autologous cell therapies, but to date most current or approved clinical trials are using fully characterized allogeneic or non-allogeneic cell banks alongside immunosuppressive drugs. Until now, all current or approved clinical trials utilize iPSC generated using EBNA1 expressing plasmids containing the *OriP* sequence to maintain a self-replicating episome. These vectors are amplified in bacterial hosts and contain bacterial DNA motifs recognized by the transfected cells innate and intrinsic interferon host defense responses. Moreover, the continued forced expression of the Epstein-Barr virus EBNA1 protein is known to cause widespread alterations in gene expression as well as elevated oxidative stress and DNA damage occurrence. Additionally, this method of iPSC derivation incorporates a short-hairpin RNA (shRNA) for the p53 protein; often referred to as the *guardian of the genome*. The shRNA functions to transiently silence the expression of p53 protein and has been demonstrated to result in an increased persistence of DNA damage in iPSC produced this way. All of these factors have significant implications for the safe clinical use of iPSC generated using *oriP*/EBNA1 plasmid episomes.

The aim of my project was to investigate the function of a novel system in reprogramming and iPSC development. Doggybone DNA (dbDNA) vectors are free of *oriP*/EBNA1 sequences, bacterial motifs and are produced in a chemically defined, low endotoxin, cGMP compliant manufacture. My results describe efficient iPSC reprogramming by applying equivalent gene sequences transiently expressed from dbDNA vectors in protocols employing both animal-derived and animal-free constituents when using weight equivalents of both systems and not molecular equivalents. In direct comparator experiments with the current state-of-the-art *gold standard oriP/*EBNA1 episomes, dbDNA vectors produced iPSC colonies with the same efficiency but dbDNA-iPSC displayed evidence of greater stability in terms of maintenance of pluripotency. Differential transcriptomic evaluations by microarray showed that the persistence of *oriP*/EBNA1 episomes resulted in an elevated interaction with *immune system processes* and *IFN signalling* in iPSC when compared to dbDNA generated iPSC. Moreover, an increased susceptibility for DNA damage incitement alongside unwanted spontaneous differentiation in iPSCs incorporating the

*oriP*-EBNA1 were all demonstrated and showed to be intrinsically linked to one-and-other. We propose a potential that utilizing dbDNA vectors presents a safer and more stable approach to iPSC production and development and that this could, with further work, help to bridge the gap between iPSCs and their greater clinical translation.

**Table of contents:**

## Table of figures:

**Acknowledgements**

Firstly, I would like to express my sincere gratitude to my director of studies Prof. Tristan McKay for his continuous support throughout my Ph.D. study. His patience, motivation, and immense knowledge has been unwavering over the 3-year period and his supervision has helped me both in the research and writing of this thesis. I could not have imagined having a better mentor and am incredibly grateful to be remaining under his guidance to begin my KTP position with MMU and Immetacyte.

I would also like to thank my second supervisor, Dr. Stephen White, for his insightful comments and encouragement over the past three.

Likewise, my thanks and appreciation go out to Touchlight Genetics for providing me with the opportunity to be able to realize a career ambition of my own in carrying out this Ph.D. study. With thanks specifically to Dr John Tite, Dr Lisa Caproni, Dr Kinga Karbowniczek and the rest of the team for their hard work and support during this time.

I would like to thank my fellow members of Team McKay: Alysha Burrows, Louise Bullen, Aseel Sharaireh, Stuart Fielding and Rhys Smith for their support and help over the last 3 years. But I would also like to express thanks to all my colleagues and now friends here at MMU for making this experience a truly enjoyable one.

My words could not cover the help, support and love expressed by my Mum and Dad during this period. They have supported me in every possible way during my studies and I will forever be indebted to them for this.

Last but not least, I would like to thank my partner, Lorna, for supporting me both in and out of a work context. As with my family, I will forever be indebted to you for everything you have done for me over this period.

# 1.0- Introduction:

## 1.1. **Classification of stemness:**

The term meristem was coined by botanist Karl Nägeli in 1858. It is still used to describe populations of self-renewing cells within plants that retain a capacity to divide, and are responsible for plant growth and organ production (Kastin, 2013). With a clear functional likeness, it has been suggested that the term *'stem cell'* was derived from its botanic counterpart (Ramalho-Santos and Willenbring, 2007; Monti et al., 2012). Stem cells are unspecialized cells of the human body that are able to perpetually self-renew in this state, whilst also retaining an ability to differentiate into all somatic cell types when exposed to appropriate stimuli.

A stem cells ability to form a spectrum of different somatic cell types is referred to as its '*potency*'. Consequently, stem cells are grouped based on their relative potent properties. Potency can be viewed hierarchically, with totipotency at the apex of the hierarchy and unipotency at the base - Figure 1. Derived from the Latin word *totipotentia* meaning an '*ability for all things'*, human totipotent stem cells are capable of forming all embryonic and extra-embryonic cell types (Mahla, 2016). Pluripotent stem cells (PSCs), derived from *pluripotential* a term meaning an '*ability for many things'*, are a step below totipotent stem cells in terms of their potency. PSCs can form all cells of the three somatic germ lineages (mesoderm, endoderm & ectoderm), and so are accountable for the formation of all cell types within the human body. However, unlike totipotent stem cells, PSCs cannot develop extra-embryonic structures such as the placenta. As stem cells commit to becoming more specialised, their differentiation capacity and relative potency becomes more limited. Multipotent stem cells can form a number of different cell types but are confined to a specific lineage - such as a haematopoietic stem cell which can only specify to form blood cells. The most restricted stem cell class in terms of its potency are unipotent stem cells. This cell type has an ability to self-renew and maintain a stem-like state, however is only capable of producing one cell type (A and B, 2011).

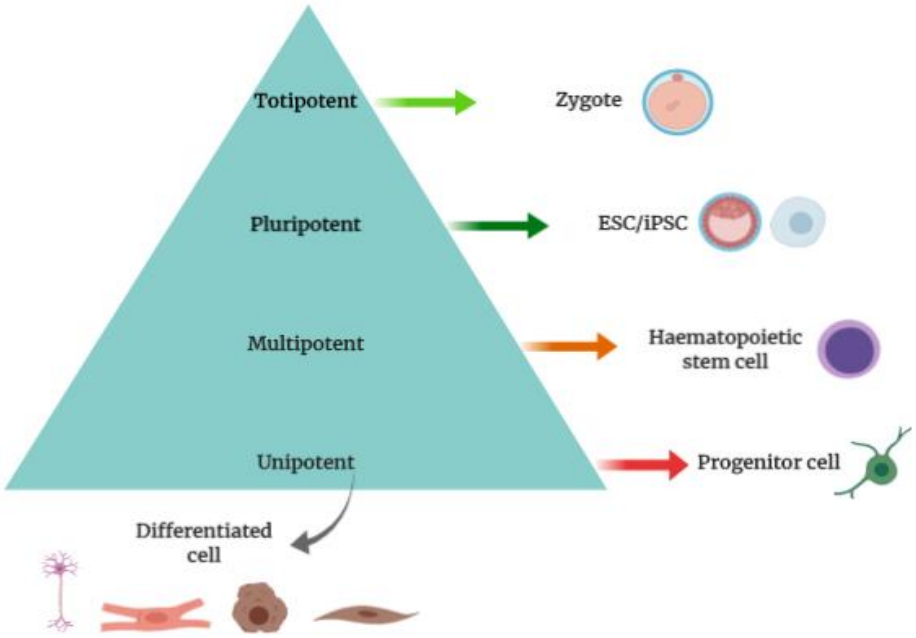



***Figure 1. Hierarchy of potency and their associated somatic cell types.***

### 1.1.1. **Stemness in early embryogenesis:**

During early human embryogenesis, potency stages can be mapped onto different phases of the developing embryo (Figure 2). A post fertilization zygote (E0-5) is the only human totipotent stem cell source. The zygote is capable of forming all cell types relating to whole organism development, including extra-embryonic structures. Then, at approximately E6, the zygote develops into a blastocyst. The blastocyst consists of the discernible trophectoderm (TE) and inner cell mass (ICM). The TE is responsible for the development of extra-embryonic support structures such as the placenta, umbilical cord and chorion (Niakan et al., 2012). The ICM retains an unspecialised state, consisting of a pluripotent cell type often termed embryonic stem cells (ESCs/ES cells). As embryogenesis progresses beyond approximately E8, ES cells from the ICM become more specialised reaching the multipotent stage and are therefore limited in terms of their potency to a single germ lineage. Much later in the developmental process, the cells will then ultimately reach unipotency before coming to a terminally differentiated state (Zakrzewski et al., 2019). This timeline provides a key insight into the transient and short-lived nature of totipotent and pluripotent stem cells in embryogenesis. Totipotent stem cells exist only within the first 5 days post-fertilisation,

whilst ESCs exist only within a ~48-hour timeframe, both pertaining to an incredibly transitory state.

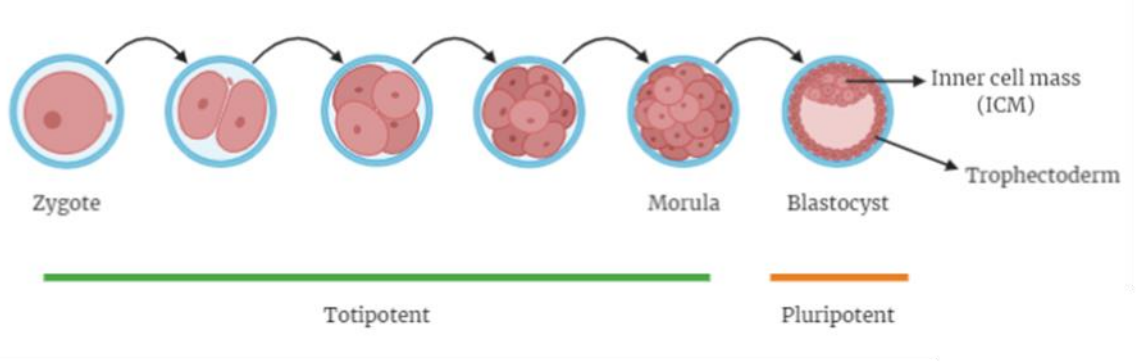



***Figure 2. Early embryogenesis up to blastocyst formation.***

### 1.1.2. **ESCs – derivation, roadblocks & recovery:**

The National Institute of Health (NIH) define human ESCs as "*cells that are derived from the inner cell mass of blastocyst stage human embryos, that are capable of dividing without differentiating for a long period in culture and are known to develop into cells and tissues of the three primary germ layers*" (*Glossary | stemcells.nih.gov*, n.d.). ES cells, and their unique pluripotent characteristics, have clear research and clinical promise. It was believed that the derivation of this cell type would herald a new modern era for science, the era of regenerative medicine. In 1981, the first successful derivation of a mouse embryonic stem cell line was carried out (Martin, 1981). This pioneering work effort, alongside exponential improvements in culturing techniques, eventually led to the isolation of the first stable human ES cell line (Thomson et al., 1998). The isolated cells were demonstrated to express key pluripotent markers (Alkaline phosphatase$^{+}$, SSEA3$^{+}$ Tra-160$^{+}$ & Tra-181$^{+}$) and possess a capacity to form teratomas expressing all three somatic germ layers in immunodeficient mice. All this to say, the successful derivation of ESCs resulted in a multitude of new prospects, ranging from cell replacement therapy for a plethora of disorders, '*off the shelf' ex-vivo* organ development, disease modelling and novel drug screening mechanisms (Vazin and Freed, 2010).

Despite steps being taken to develop the methodology of the derivation and culturing of ESCs, there was critical opprobrium generated by the field of stem cell research vis-à-vis the inviolability of life. Early methods of ESC isolation obligated the destruction of an embryo. As such, moral and ethical complications soon became entangled within stem cell research. The most vociferous of arguments arose from religious communities in relation to the disregard for the sanctity of life that is central to providing the starting material to effectuate this line of research. The ethical volatility surrounding ESC derivation soon engendered global political engagement (Figure 3). The Dickey-Wicker amendment and the subsequent prohibition of federal funds by President Bush stymied any research with embryonic stem cells across the U.S.A. During this hiatus, new players entered the stage of pluripotency. Yet, embryo destruction, the raison d'etre for stymying stem cell research, was soon negated with the isolation of ESCs from a single blastomere (Chung et al., 2008). This, alongside a reversal of funding prohibitions has exponentially increased ESCs therapeutic potential. Over 20 clinical trials using ESC-derived cells are ongoing, targeting conditions such as macular degeneration (NCT01344993), type 1 diabetes (NCT03162926) and Parkinson's disease (NCT03119636) (Eguizabal et al., 2019).

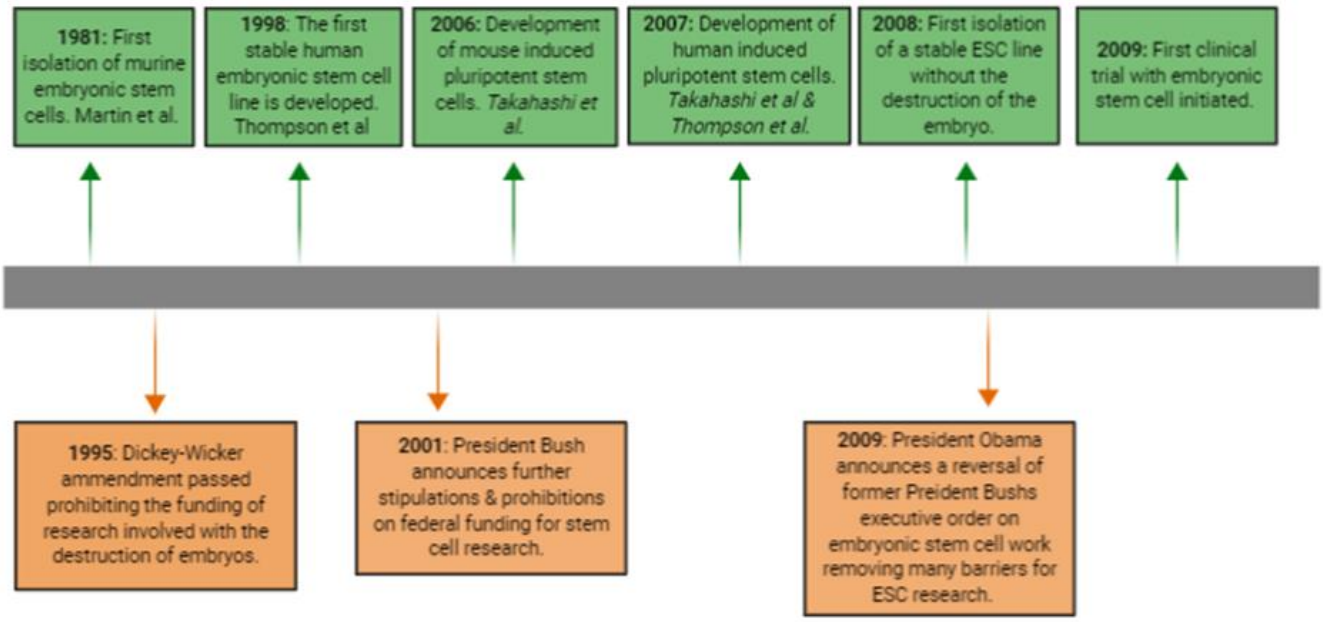



***Figure 3 - Progression of embryonic stem cell research in parallel with American research sanctions.***

***Table 1 - All trials, recruiting, ongoing and completed utilising ESCs/ESC-derived cell***

| No. | Cell origin | Cell-type produced | Country | Status |
|---|---|---|---|---|
| NCT01217008 | hESC | Oligodendrocytes | USA | Completed |
| NCT01344993 | hESC | Retinal pigment epithelium (RPE) | USA | Completed |
| NCT01345006 | hESC | RPE | USA | Completed |
| NCT01469832 | hESC | RPE | UK | Completed |
| NCT01625559 | hESC | RPE | Korea | Unknown |
| NCT01674829 | hESC | RPE | Korea | Unknown |
| NCT01691261 | hESC | RPE | UK | Ongoing |
| NCT02057900 | hESC | CD15+Isl-1+progen | France | Recruiting |
| NCT02239354 | hESC | β-Cell progenitors | USA | Ongoing |
| NCT02286089 | hESC | RPE | Israel | Recruiting |
| NCT02302157 | hESC | Oligodendrocyte progenitors | USA | Recruiting |
| NCT02445612 | hESC | RPE | USA | Ongoing |
| NCT02452723 | phESC | Neural stem cells | Australia | Recruiting |
| NCT02463344 | hESC | RPE | USA | Ongoing |
| NCT02590692 | hESC | RPE | USA | Recruiting |
| NCT02749734 | hESC | RPE | China | Recruiting |
| NCT02755428 | hESC | RPE | China | Recruiting |
| NCT02903576 | hESC | RPE | Brazil | Recruiting |
| NCT02941991 | hESC | RPE | UK | Ongoing |
| NCT03046407 | hESC | RPE | China | Recruiting |
| NCT03119636 | hESC | Neural Precursors | China | Recruiting |
| NCT03162926 | hESC | β-Cell progenitors | USA | Recruiting |
| NCT03163511 | hESC | β-Cell progenitors | USA | Recruiting |
| ChiCTR-OCB-1500596 | hESC | Corneal epithelium | China | Recruiting |
| ChiCTR-OCB-15007054 | hESC | RPE | China | Ongoing |
| ChiCTR-OCB-15007055 | hESC | RPE | China | Ongoing |

## 1.2. **Induced pluripotent stem cells:**

### 1.2.1. **History of iPSC development:**

August Weismann devised a genetic theory of heredity at the end of the 19th century commonly known as the Weismann barrier. He postulated that for cells to become committed to a specific, differentiated state, any unnecessary genetic code must be removed or terminally inactivated (Weismann, A. 1893). In the 20th century, Conrad Waddington developed the, now widely renowned, Waddington's landscape theory. This, of a similar ilk to Weismann's theory, suggests that cell fate and differentiation is a perpetual and irreversible state. Waddington's landscape depicts cellular differentiation as a ball at the top of a hill (*Figure 4*). As the ball descends and begins rolling down the hill, this metaphorically represents the cell transitioning from its pluripotent/precursor state to a differentiated state, with the cell ultimately reaching its terminal fate at the bottom of the hill (Waddington, 1957). Both theories, although originating from different lines of research both represent the idea that cellular differentiation is an irreversible, unidirectional process.

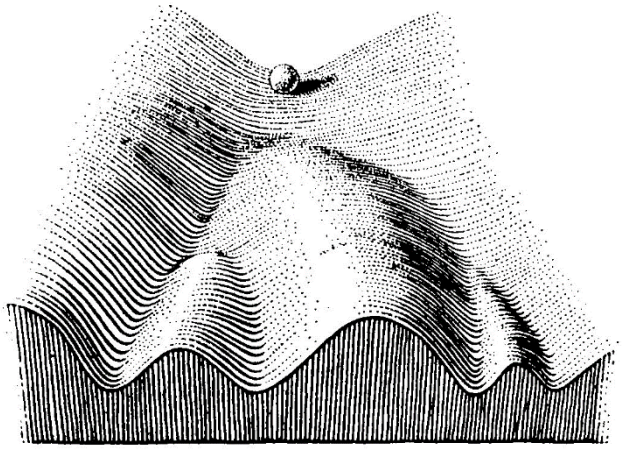


***Figure 4 - Waddington's landscape: image depicting a ball at the top of a hill to represent embryonic cell differentiation as the ball progresses down the hill to its terminally differentiated state. The crevices in the hill represent different potential fate outcomes for the cell dependent on its path (Source: Waddington, 1957).***

Somatic cell nuclear transfer (SCNT) is the process of isolating a nuclei from one cell type and transplanting it into a second enucleated cell (Briggs and King, 1952). The derivation of this technique was the basis for the innovative work carried out by Sir John Gurdon. Gurdon utilised SCNT to transfer the nucleus from a differentiated intestinal tadpole cell to an enucleated recipient oocyte. The oocyte, containing the transplanted nucleus, was able to develop into a healthy feeding tadpole. Thus, Gurdon demonstrated that a nucleus, which can promote the formation of a differentiated cell, can simultaneously retain the genetic information necessary for the formation of all other cell types too (Gurdon, 1962). This seminal work permitted an important and novel insight; development and differentiation are not irrevocable changes (Stadtfeld and Hochedlinger, 2010). The birthing of a number of different mammals, most notably Dolly the sheep, adopting the use of SCNT and nuclei from terminally differentiated cell types, again corroborated the theory initially developed by Gurdon (Wilmut et al., 1997; Takahashi and Yamanaka, 2016).

It was first demonstrated in the 1980's that the ectopic overexpression of lineage associated transcription factors can differentiate one cell fate directly to another, a process termed *transdifferentiation*. The overexpression of *MyoD,* a transcription factor associated with myogenesis and muscle development, was found to convert terminally differentiated fibroblast cells into myoblasts (Davis et al., 1987). This premise was then adopted by several labs alluding to the transdifferentiation of a broad range of different cell types. The body of work carried out provided unequivocal evidence that mature cells can differentiate across and within germ lineages to different mature cell types; without reverting to a stem cell state (Stadtfeld and Hochedlinger, 2010).

In 2001, work from Takashi Tada's lab in Japan focused on the production of a fusion cell type, with the foundations of the process being outlined for over a decade (Blau et al., 1983). The lab fused T-cell progenitors, thymocytes, with ES cells mediated by electric fusion. The resulting cell type demonstrated functional and transcriptional characteristics of the reactivation of inactive X chromosomes, and were part epigenetically reset to the point of endogenous pluripotency gene expression (Tada et al., 2001). The derivation of cell fusion models demonstrated a possibility of resetting somatic cell types to a pluripotent status, and, what's more, that transcription factors exist that may help to mediate this process (Takahashi and Yamanaka, 2016).

In the early 2000s, research with ESCs, as aforementioned, became restricted in some areas as a result of ethical and moral complications. Scientists working within the field looked to bypass the complications of ESC use, whilst still retaining cells with a pluripotent capacity and vast research potential. These seminal, yet seemingly separate pieces of work would ultimately provide the basis for a break in the PSC research stalemate. A search for transcriptional regulators of pluripotency, that if identified and over-expressed, could potentially convert a terminally differentiated cell type to a pluripotent phenotype was underway.

It was long established that a transcriptional hierarchy exists which is critical to the specification of ES cell identity as well as the cells unique functional properties. *OCT4 (POU5F1)* has been identified as a protein critical to the maintenance of pluripotency, as its loss of function in ES cells results in their unwanted differentiation (Nichols et al., 1998). *OCT4* is a member of the POU class of homeodomain proteins, a heavily conserved set of proteins with functional attributes within cell fate specification (Boyer et al., 2005). Along with *OCT4*, *NANOG* and *SOX2* are also central to the transcriptional hierarchy necessary for pluripotency maintenance. It is now clear that these three factors function in-conjunction with one and other to control a large population of downstream regulatory genes. This is achieved by the protein's co-occupancy of promoters for other developmentally important homeodomain transcription factors, permitting the maintenance of the specialised ESC regulatory network. With a developed understanding of the main transcriptional regulators of pluripotency, this would be key in narrowing the search for factors critical to the de-differentiation of somatic cells back to pluripotency.

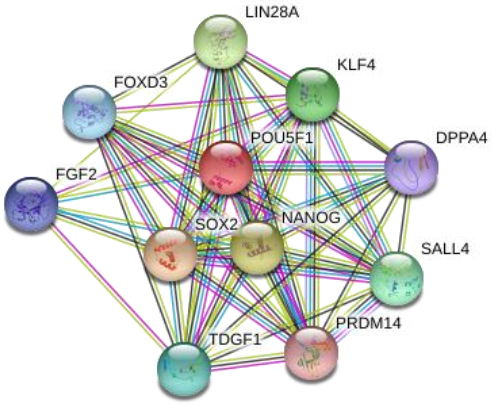



***Figure 5 - Pluripotency network: transcription factors necessary for the maintenance of pluripotency. OCT4 (POU5F1), SOX2 and NANOG are critical to this process and regulate the transcription of several different genes permitting the maintenance of pluripotency.***

In an initial search for *Oct3/4* target genes, work by the Shinya Yamanaka lab isolated F-box containing protein (*Fbx15*) as a novel target of *Oct3/4.* It was well established that the expression of *Oct3/4* in ES cells was critical for self-renewal (Nichols et al., 1998). The Yamanaka lab was subsequently able to correlate a direct relationship between *Oct3/4* inactivation and the extinction of *Fbx15* expression too. However, despite this specific relationship, the loss of *Fbx15* did not promote any developmental defects in homozygous mutant mice and thus was concluded to be dispensable in terms of the maintenance of ES cell self-renewal (Tokuzawa et al., 2003). The Yamanaka lab subsequently wanted to elucidate if a panel of 24 candidate genes would be able to induce a pluripotent phenotype in terminally differentiated cells and if so, which transcription factors would be critical to this process. This involved the development of an assay in which the induction of pluripotency, if achieved, would be detected as a resistance to G418 (Geneticin). The genes responsible for this antibiotic resistance were inserted into the prior mentioned *Fbx15* gene via homologous recombination. This would provide an ideal marker as the *Fbx15* gene is upregulated specifically in murine ES cells and early development but as mentioned is dispensable in relation to the maintenance of pluripotency. The retroviral transduction of all 24 candidate genes into mouse embryonic and adult fibroblasts gave rise to G418 drug-resistant colonies, with morphologies similar to that of ES cells. The cells produced were termed *induced pluripotent stem cells* (iPSCs/iPS cells). Subsequently, each of the 24 factors were removed individually from the remaining 23 and the effect on

colony formation after 10 and 16 days was noted. This allowed the lab to isolate 10 factors which they continued to compare in the same manner before a final 4 factors were isolated as being important in the generation of induced pluripotent stem cells – *Oct3/4*, *Sox2*, *Klf4* & *c-Myc* (OSKM). Figure 6 summarises the idea that development and differentiation is not a linear and irreversible process, using a modified version of Waddington's landscape.

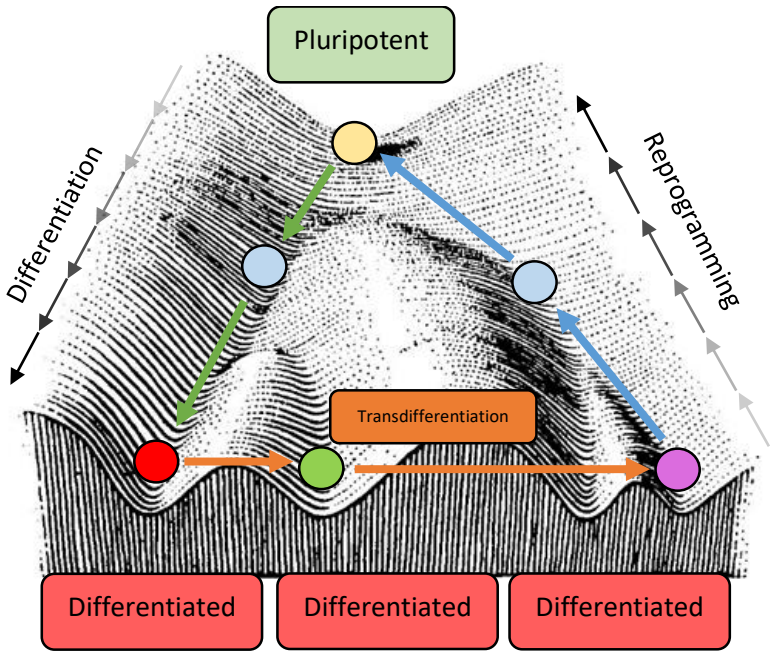



***Figure 6 – Waddington's landscape re-examined: adopted version of Waddington's landscape to include the updated ideals of transdifferentiation and reprogramming.***

The development of human iPSCs followed within a year of producing the murine counterparts. Adult human skin fibroblasts were successfully reprogrammed back to a pluripotent state using human equivalents of the same OSKM transcripts that successfully produced the first mouse iPSC line (Takahashi et al., 2007). Concurrently, a lab led by James Thompson also successfully produced human fibroblast-derived iPSCs (Yu et al., 2007). The Thompson lab however, did use a different combination of human transcription factors – maintaining *OCT3/4* and *SOX2* while replacing *KLF4* and *c-Myc* with *NANOG* & *LIN28*. A multitude of subsequent papers and laboratories were able to reproduce this same process and procure iPSCs using both OSKM and OSNL factors with similar efficiencies. In both the Yamanaka and Thompson lab, viral transduction was utilised as the vehicle or vector to deliver the pluripotency factors to fibroblasts (Figure 7). Yamanaka utilised a retroviral delivery system, whilst

Thompson, a lentiviral system, both with the aim of having prolonged ectopic transgenic expression of pluripotency transcripts. The prolonged expression would ensure complete induction of endogenous pluripotency gene expression and ultimately, the development of *bona-fide* iPSCs. Both iPSC types were demonstrated to have typical ESC-like structure (round morphology, scant cytoplasm and large nucleoli) and express key endogenous pluripotency markers (*SSEA3*, *SSEA4*, *TRA-1-60*, *TRA-1-81*, *NANOG*). Likewise, the cells were demonstrated to maintain a capacity to form cells of all 3 germ lineages too.

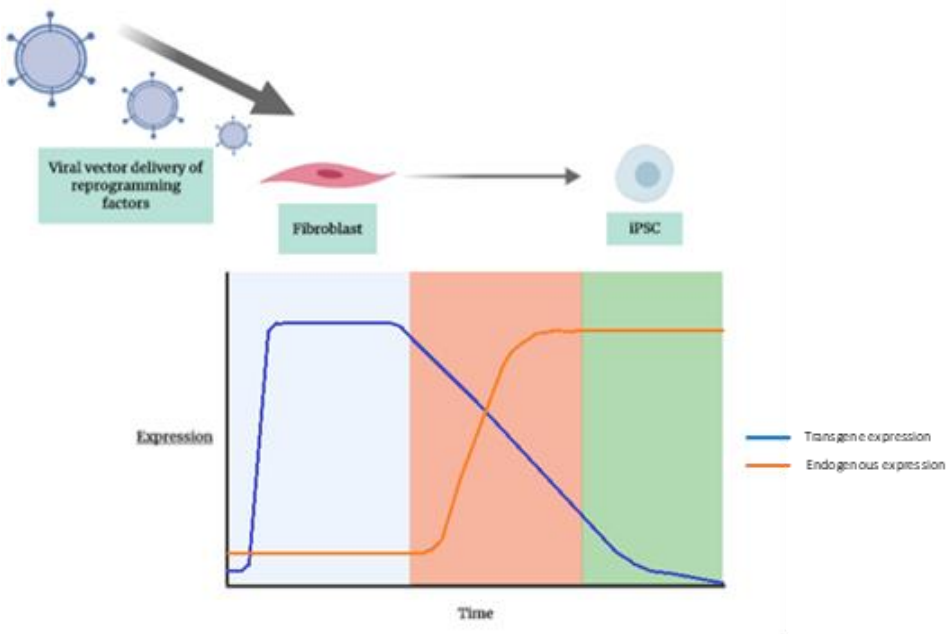



***Figure 7 - Process of reprogramming somatic cells to iPSCs. Depiction of prolonged transgene expression which ultimately drives endogenous pluripotency gene expression.***

### 1.2.2. **Reprogramming:**

#### 1.2.2.1. **Phases of reprogramming**

The process of reverting a terminally differentiated cell back to a pluripotent state is termed induced pluripotent stem cell *reprogramming*. The reprogramming of differentiated fibroblasts back to iPSCs offers, as the use of ESCs does, a platform to continue to develop the potential for regenerative medicine as well as studying the mechanisms of early development. However, with the possibility of deriving the somatic cellular material (dermal fibroblasts) necessary to produce iPSCs in an autologous fashion, this mitigates any immunological complications previously identified with ESC use. Moreover, being derived from adult somatic cells, this may help to assuage previous ethical and moral complications surrounding the isolation and use of ES cells.

The reversal of development and epigenetic erasure that ensues when reprogramming terminally differentiated somatic cells is a complex process. Sustained transgene expression is necessary to facilitate chromatin re-arrangement on both a global and local scale – ultimately mediating the silencing of lineage specific genes and an activation of pluripotency-related ones (Biran and Meshorer, 2012). The exogenous over-expression of key pluripotency transgenes has also been demonstrated to result in the ordered progression of X-chromosome reactivation (Xa) in both murine and human iPS cells (Cantone and Fisher, 2017). In humans cells, the reactivation of the X-chromosome is however unstable. The reprogramming process, despite facilitating Xa, doesn't yield wholly reactivated iPSC colonies - with culturing conditions being implicated in such variability (Kim et al., 2014). The process of reprogramming remains, on the whole, elusive and specifically in its early stages, largely stochastic in nature (Buganim et al., 2012).

Upon the successful derivation of iPSCs from both murine and human cells, work had been undertaken to elucidate the mechanism behind the global cellular resetting back to pluripotency. Extensive time course transcriptomic analysis permitted great insight into the process of reprogramming and its subsequent grouping into 3 main phases: initiation, maturation and stabilization (Samavarchi-Tehrani et al., 2010).

There are a number of different fates which may be exhibited by cells in the initiation phase of pluripotency onset, such as apoptosis, senescence, transdifferentiation and reprogramming (Buganim et al., 2012). Of cells that undergo faithful reprogramming, it has been elucidated that all of these cells must acquire a number of mesenchymal to epithelial transition (MET) related modifications. This ubiquitous nature of MET-onset during reprogramming was demonstrated by backtracking Oct4/Nanog positive iPS cells using single cell time lapse microscopy (Smith et al., 2010). The initiation phase is therefore characterised largely by the cells undergoing MET. The process of MET instigates changes to the cellular transcriptome, with a loss of somatic cell signatures and an upregulation of epithelial related genes. Moreover, MET also initiates changes to cellular morphology, with this morphological re-construction alone, being demonstrated to hold the capacity to initiate epigenetic changes associated with pluripotency regulation (Downing et al., 2013). Subsequently, reprogramming cells undergoing MET also develop ESC-like features, including an increased proliferation alongside an increased resistance to apoptosis and senescence (Hong et al., 2009). It has become apparent that despite the probabilistic nature of the initiation process, it occurs with a degree of systematism, with the requirement being that cells hit well-defined MET checkpoints as opposed to the synchronicity in which they do so (Buganim et al., 2013; David and Polo, 2014).

Predominantly, the changes associated with the maturation phase of reprogramming are, on the whole, well-established. This phase, and its transitioning cells, are largely defined by major transcriptional changes, specifically the onset of pluripotency-related genes (Buganim et al., 2013; David and Polo, 2014). However, the initial period of the maturation phase has been determined to be largely random in nature. Exogenous expression of the Yamanaka factors (OSKM), ultimately gives rise to a somewhat stochastic upregulation of pluripotency transcriptional regulators whose specific order remains mostly unknown. Yet, beyond the stochastic activation of said genes, in the latter maturation phase, the transcriptional activation becomes more deterministic. The core pluripotency genes such as OCT4, SOX2 and NANOG will all be readily upregulated during this period, along with their downstream targets. This suggests that, despite the stochastic nature of transcriptional onset during the early maturation phase, that a hierarchy of pluripotency gene activation exists towards its latter end and is critical to the development of fully-fledged iPSC formation.

The final stage of iPSC development is the stabilisation phase. At this point the cells have become iPSCs – expressing key pluripotent markers and having typical morphological characteristics. The stabilisation phase encompasses an assessment of the newly-developed iPSCs pluripotent signature via their ability to maintain the expression of pluripotency-related genes independent from ectopic transgene expression. At this point the cells will be passaged, expanded and subsequently tested for transgene independent endogenous pluripotency gene expression. Despite the seemingly fully-fledged, committed nature of these cells at the stabilisation phase, epigenetic alterations will continually occur during this phase and for a number of passages beyond this point too.

#### 1.2.2.2. **Transgene modifications:**

Despite an incomplete understanding of the reprogramming process, many steps have been taken to attempt to increase its efficiency, as well as increasing the number of different somatic cell types that can be manipulated back to pluripotency. One area of reprogramming susceptible to manipulation has been the use of different pluripotency transgenes. Since its initial description in humans in 2007, the OCT4, SOX2, KLF4, c-MYC transcription factor cocktail described by the Yamanaka lab has been a mainstay of the reprogramming protocol, being more readily employed than other published counterparts. Since then, a number of different transgenes have been tried and tested in synergistic combinations, but none seemingly long-standing enough to challenge the *status quo*.

Yet, published in 2011, the Yamanaka lab outlined how the inclusion of a short-hairpin RNA for p53 (p53-shRNA) and replacing c-Myc for L-Myc helped improve the quality of the iPSCs produced (Okita et al., 2011). L-Myc is, like c-Myc, a member of the *myc* oncoprotein family. Yet, L-Myc is reported to be much less oncogenic, being far less associated with cancer development in comparison to the more volatile c-Myc (Nesbit et al., 1999). The Okita lab reported that the addition of these two factors increased reprogramming on average by approximately 3-fold in comparison to the highest efficiency achieved by a combination of any other reprogramming factors. It is a well-established line of thought that p53 is an inhibitor of successful reprogramming. Often referred to as the *guardian of the genome*, p53 is responsible for a number of processes such as senescence, apoptosis and DNA damage response (Horikawa et

al., 2017). The functional properties asserted by p53, do therefore make its function incompatible with the self-renewing properties of iPSCs, further perpetuating the role of p53 as a barrier to iPSC production (Zhao and Xu, 2010). The inhibition of p53 expression, by blocking mRNA translation, has been demonstrated to dramatically increase the efficiency of reprogramming and murine iPSC development for over a decade (Hong et al., 2009). That said, the inclusion of further genetic modifications permits more implications for the downstream use of iPS cells in the clinic and beyond. Loss of wild-type p53 function is a prerequisite to the onset of the majority of human cancers. With p53 being responsible for the suppression of genomic instability, its transcriptomic inactivation and subsequent reduced functionality increases the possibility of genomic aneuploidy amongst other chromosomal aberrations (Zhao and Xu, 2010). The inhibition of p53 function during reprogramming therefore poses considerable concerns about the tumorgenicity and genomic integrity of the iPSCs, with identifiable chromosomal aberrations being commonly reported (Chin et al., 2009; Hong et al., 2009). Moreover, an extensive study characterising 140 independent ESC lines was carried out using whole exome sequencing (WES) which isolated a number of previously undetected mutations to the *TP53* gene. Such affected residues in the ESCs were likewise commonly mutated in human tumours. The mutational abundance on each allele was then quantified using droplet digital PCR (ddPCR). Such cells were therefore demonstrated to naturally acquire p53-related mutations as commonly seen in human cancers and moreover, that such mutated allelic fractions increased in proportion to the cells passage number. This suggests that the p53 mutation confers a selective bias which may go unnoticed in most applications of such cells, with the paper highlighting the necessity of careful genetic analysis of any prospective clinical application of such cells (Merkle et al., 2017). Despite the suggestion of its rate-limiting functionality on the reprogramming process, the suppression of p53 expression is not a necessity for the success of somatic cell reprogramming to iPSCs. However, suppression of its function is well-characterised to result in cancerous transformation of cells which render then clinically redundant. Therefore, a clear rationale exists to actually allow the natural persistence of p53 during the reprogramming process. Moreover, work from the Ichida lab demonstrated how an inherent antagonism between transcription and proliferation is a limiting factor in cellular reprogramming. Transcription factor overexpression during reprogramming within hyperproliferative

cells is critical to lineage conversion. However, sustenance of transgene expression throughout the reprogramming process impedes cell proliferation and DNA replication – processes necessary for its completion (Babos et al., 2019). However, natural populations of simultaneously hypertranscribing, hyperproliferative cells (HHCs) exists. Furthermore, it has been suggested that the expression of wild-type p53 increases the population of the privileged HHC cell types in comparison to p53 knock-down cells during reprogramming. WT-p53 is suggested to interact with topoisomerase related genes and that a loss of p53 function reduces nuclear expression of these genes and results in a reduction in HHC levels. This newfound data suggests the role of p53 is more nuanced than first thought in the context of cellular reprogramming (Babos et al., 2019).

Duly, the process of increasing the number of transgenes to improve the efficiency of reprogramming may or may not improve the clinical applicability of iPSCs. As prior mentioned, the mechanism of reprogramming remains incompletely understood, therefore the true implications of added transgenic material cannot be unequivocally elucidated. However, in terms of the applicability of iPSCs into the clinic, an argument exists that the more simplistic route with minimal genetic modification, may in turn become the most successful.

#### 1.2.2.3. **Vectorology:**

Another area that has received considerable research effort and is a critical factor to the reprogramming process, is the type of transgene delivery system used. The transgene delivery vehicle is referred to as the *vector*. The choice of vector when reprogramming somatic cells is a pivotal decision affecting both the efficiency (*number of iPSC colonies produced in respect to the number of fibroblasts subjected to reprogramming*) and quality of iPSC developed, which in turn is also determinative of the cells prospective downstream use.

When first derived from human fibroblasts in seminal studies in 2007, iPSCs were generated using genome integrating viral vectors in both the Yamanaka & Thompson labs (Takahashi et al., 2007; Yu et al., 2007). The Yamanaka lab utilised a γ-retroviral vector to develop human iPSCs in a proof-of-principle study. This method of iPSC

production is robust and pertinent in this respect. However, permanent modifications to the genome of the transduced somatic cell makes the subsequent iPSCs unsuitable for clinical use. Takahashi reported that each clonal iPSC line generated had between three and six retroviral integrations per reprogramming factor. This meant that every clone therefore had more than 20 retroviral integration sites in total (Takahashi et al., 2007). The permanent and semi-random nature of such integrations of transgenic DNA into a somatic cell genome can increase the risk of tumorigenesis via insertional mutagenesis. The retroviral vector will, by nature, integrate within the host cell genome. Preferences of such semi-random insertions lie between actively transcribed genes and *cis*-regulatory elements (*CpG* motifs) (Baum, 2007). Dependent upon the site of the insertion, such transduced cells may develop an induced, serial replication capacity providing them with a clonal bias. Such clonal imbalance and subsequent clinical transfer of these cells into a patient has, in previous studies using transduced haematopoietic cell types, resulted in the onset of Leukaemia or sarcoma (Hacein-Bey-Abina et al., 2003). Hence, there is a clear prospect that utilising an integrating vector can result in a secondary morbidity. This was re-iterated in murine studies where approximately 20% of mice derived from iPSCs reprogrammed using the γ-retroviral vector developed tumours. The aetiology of the tumours were, in part, determined to be associated with the reactivation of c-Myc retroviral DNA (Okita et al., 2007).

It was recognised that for iPS cell therapies to progress into the clinic, that alternative reprogramming methods had to be developed that negated any integration into the manipulated somatic cell genome. As such, considerable research efforts have been applied to the area of non-integrating vectorology in an attempt to diversify and improve safe methods of pluripotency protein upregulation to aid in iPSC clinical development. Now, a myriad of different non-integrating vectors have been identified and demonstrated, with varying levels of success, to be able to induce pluripotency onset and iPSC development. More recently, a multitude of small molecules and the CRISPR-cas9 technology has been demonstrated to be capable of iPSC production (Weltner et al., 2018). However, virus, plasmid DNA, mRNA and recombinant proteins are more commonly manipulated and utilised for iPSC development. These vector types have long-been identified and as such, much progress has been made to develop their place within the iPSC research field.

#### 1.2.2.3.1. **Sendai virus (SeV)**

Belonging to the Paramyxoviridae family of viruses, SeV is an enveloped virus having a non-segmented, negative-sense, single-stranded RNA genome. SeV is non-pathogenic in humans and will typically complete its cell cycle in the infected host cells cytoplasm, with its viral genome encoding a total of six critical proteins (Nakanishi and Otsu, 2012). The Hasegawa lab was first to report the production of iPSCs using a SeV vector (Fusaki et al., 2009). The report details the use of a fusion protein-deficient SeV vector to reprogram and produce *bona-fide* iPSCs. The vector reprogrammed human dermal fibroblasts at efficiencies from 0.001-1%, in a non-integrative manner. Transgene expression is reduced proportionately with cell division, and this loss was reported to be accelerated by the addition of antibodies to the haemagglutinin (HN) surface envelope protein, small interfering RNAs to RNA-dependant RNA polymerases (RdRp) or by modifying the SeV to be temperature sensitive (Seki et al., 2010; Ban et al., 2011). The production of SeV vectors ultimately apexes with the generation of the replication-deficient, auto-erasable SeV (Nishimura et al., 2017). The vector is designed to self-regulate its expression in relation to the presence of specific micro-RNAs/miRNAs. It has long been recognised that with the inclusion of miRNA target sequences in the 3'UTR (untranslated region) of a gene, that its expression will be downregulated in the presence of its targeted miRNA (Brown et al., 2007). Subsequently, a SeV vector was produced with target sequences to miR-302 within the 3'UTR of the large protein (L) gene within the SeV genome. The expression of miR-302 is ubiquitously upregulated in pluripotent cell types, with diminished expression in terminally differentiated cells. Subsequently, within a differentiated cell, the SeV will persist and express its exogenous pluripotent transgenes. As the differentiated cells reprogram and develop a pluripotent state, expression levels of miR-302 will increase which will inversely promote a decrease in SeV levels. This process should therefore result in integration-free, SeV-free iPSCs with minimal effort (Nishimura et al., 2017; Borgohain et al., 2019)

Duly, there are clear advantages to employing the SeV system within the process of reprogramming and iPSC production. The virus, by nature, binds to the ubiquitously expressed, membranous, sialic acid receptor and as such proffers the virus with a wide-ranging tropism (Nakanishi and Otsu, 2012). The vectors tropism is reflected

within its ability to efficiently infect and reprogram a broad range of differentiated cell types, including: peripheral blood, fibroblasts, urine-derived cells and keratinocytes (Yang, 2014; Cristo et al., 2017; Boonkaew et al., 2018). The viral vectors transgenic requirements can be easily manipulated to ensure competitive reprogramming rates, and its lack of a DNA phase reduces the SeVs susceptibility to any silencing or modifications engendered by the infected host cell (Borgohain et al., 2019).

That said, drawbacks exist with regards to the use of the SeV system and how this can negatively impact iPS cells and their path to the clinic. Firstly, the process of SeV production is laborious and technically challenging in comparison to the production of other viral delivery systems (Rao and Malik, 2012). There are commercially available kits to combat this issue. Yet, these kits are extremely costly, this therefore limits the prospect of a homogeneous, replicable global method of iPSC production. This makes the SeV system unsustainable in terms of developing iPSCs for clinical appliances and also deters the potential for autologous cell therapies. That said, concerns also exist over the SeV vectors cytotoxic and immunogenic side-effects too. The Chen lab demonstrated that Sendai potentiates severe cytotoxic side-effects on cells undergoing transduction. The virus therefore had to be batch tested and an MOI titration carried out to perpetually analyse and minimise the vectors potential negative cellular impact (Beers et al., 2015). Moreover, despite newer modified SeV vector types being deficient of fusion, matrix and hemagglutinin-neuraminidase proteins, both fusogenic and immunogenic concerns still persist with the vectors use. Likewise, despite the vector also being non-integrative, it is common practise that all iPSCs produced by SeV undergo thorough screening protocols for integrations and the presence of viral genes - adding more complication and cost to the process (Borgohain et al., 2019).

#### 1.2.3.3.2. **mRNA:**

The transfection of reprogramming factors in the form of messenger RNA (mRNA), offers a transient and simplistic method for iPSC production. Despite descriptions of synthetic mRNA synthesis originating from the 1980s, it was not a widely utilised scientific tool, as virus, plasmid and small interfering RNAs (siRNAs) were. In the early

2000s, reports of long-lived and efficiently expressed mRNAs were published with the transcripts aiming to mimic natural mRNA features and structure (Karikó et al., 1999, 2008; Jemielity et al., 2003; Mockey et al., 2006). Typically, these mRNAs, contained 5' cap, polyA tails and untranslated regions (UTRs) as natural mRNA would, and could moreover be mass produced and delivered to different cultured cell types. The first publication surrounding the use of mRNA transfection for the reprogramming of fibroblasts and production of iPSCs came in 2010 (Yakubov et al., 2010). Despite providing evidence supporting pluripotency transgene and protein expression during the reprogramming process (OCT4, SOX2, NANOG, SSEA4), and iPSC colonies produced staining positive again for pluripotency markers (AP & NANOG), there was a degree of uncertainty surrounding the fully-fledged nature of these iPSCs (Warren and Lin, 2019). This, and subsequent attempts made to produce iPSCs using mRNA did highlight some of the key issues associated with this method of pluripotency activation. mRNA expression is extremely transient, with even the long-lived mRNAs expressing in the host cytoplasm for only 12-24 hours before undergoing degradation. Such minimal transgenic expression ultimately led to the curtailment of pluripotency protein expression and subsequently restricted the capacity for pluripotency induction and complete reprogramming within differentiated cell types. To overcome this, repeated daily transfections were undertaken to help maintain transgenic and ultimately the expression of pluripotency-related proteins. Yet, the continued, daily transfection of mRNA into cells is not only labour-intensive and time-consuming for the lab user, but repeatedly subjecting cells to transfection reagents can also pose a degree of cellular toxicity. Moreover, it has been documented that the transfection of synthetic mRNA into a host cell, activates somatic innate immune responses. Transfection of the single-stranded mRNAs activates host anti-viral responses, subsequently upregulating type-1 interferon expression (Angel and Yanik, 2010). This likewise, makes the repeated, daily transfections unsustainable again, due to an increased cytotoxicity. Continued endeavour and optimisation of this process, however, has led to the successful derivation of iPSCs from fibroblasts using both feeder and feeder-free approaches. Methods to combat innate stimulation have been adopted, such as the inclusion of B18R, a decoy receptor for type 1 interferons, which can be included as a media additive to slow the onset of the host cell inflammatory response (Warren et al., 2010). Moreover, continual adaptations of the synthesised

mRNAs has shown that the replacement of canonical DNA bases with non-normal, pseudo-nucleosides permits a greater evasion of the cell's immune response (Karikó et al., 2008). These changes have helped to permit the daily transfections of the mRNA pluripotency transgenes with less-severe repercussions. This has ultimately led to the production of iPSCs, which display a high efficiency, alongside low aneuploidy rates and limited major alterations to copy number variants (CNVs) in comparison to other reprogramming vector types (Boonkaew et al., 2018). However, the time-consuming nature of the protocol, as aforementioned, does prevent the mRNA vector from being more frequently adopted as standard practise in labs globally.

1.2.2.3.2. ***miRNAs:***

MicroRNAs (miRNAs) are short, non-coding RNAs that are approximately 22 nucleotides in length. They function in the post-transcriptional regulation of gene expression by interacting with transcribed mRNA structures based upon complimentary '*seed sequences*', usually present in the mRNA's UTR. The miRNA can then promote mRNA degradation or an inhibition of the mRNAs translation to limit gene expression and protein production. Thus through the repression of lineage-specific genes, miRNAs have been demonstrated to be able to reprogram differentiated cells back to pluripotency (Anokye-Danso et al., 2012). miRNA clusters have been isolated and identified to be able to produce iPSCs, such as miR 290-295 and miR 302-367. The latter cluster has also been demonstrated to reprogram and produce iPSCs independently, being able to replace the traditional OSKM transgenes. Integrative miRs are capable of complete reprogramming, being delivered using lentiviral vectors, but this is counter-intuitive in terms of progressing iPSCs to the clinic. Transient transfection of miRNAs has also been demonstrated to produce iPSCs, but as with the mRNA system, miRs are extremely transient and as such multiple transfections are required, which dramatically increases the workload of the protocol. Moreover, the system, although functional in producing iPSCs, does so at a low efficiency (~0.01%), making it a less-attractive option for widespread iPSC production (Miyoshi et al., 2011).

1.2.2.3.3. **Recombinant proteins:**

The isolation of pure bioactive forms of reprogramming proteins represents a novel, safe and transgene-free approach to iPSC generation. The proteins can be produced in large quantities, by manipulating prokaryotic or eukaryotic systems, and fused with cell-penetrating peptides to help facilitate its passage into the host cell. This system was first successfully utilised in murine iPSC production in 2009 (Zhou et al., 2009). The OSKM proteins produced were fused with poly-arginine PTD (11-R) at the protein C-terminus to aid cellular penetration, before being amplified in bacteria. During reprogramming, the cell medium was also supplemented with 1mM Valproic Acid (VPA), a histone deacetylase inhibitor, to attempt to increase the reprogramming efficiency (Rao and Malik, 2012). The protein was transduced into mouse embryonic fibroblasts (MEFs) containing an Oct4-GFP reporter. The kinetics and efficiency of the reprogramming was, however, considerably poorer in comparison to other vector-types. The system functioned at an overall efficiency of ~0.006%, with colonies appearing between 30-35 days after transfection. The biggest issue however, was that the transient transfection of protein alone was not sufficient to reprogram and produce iPS cells, VPA was required for the induction of OCT4 positive colonies. This far-reduced efficiency was largely in relation to the endosomal entrapment of reprogramming proteins within the host cell, rendering them useless (Zhou et al., 2009). That said, in the same year, a team in Harvard medical school were the first lab to successfully recapitulate this process in human cells, producing human iPSCs mediated by the delivery of OSKM proteins to neonatal fibroblasts. The four reprogramming proteins were fused with poly-arginine PTD (9R), again to facilitate the proteins crossing of the fibroblast cell membrane. HEK293 lines were then generated expressing the modified versions of the OSKM proteins before fibroblasts were subjected to the HEK293 protein lysates. The process, overall, was successful, leading to the derivation of iPSCs expressing endogenous pluripotency markers whilst also retaining a tri-lineage differentiation potential. That said, the protocol was also fairly arduous, with 6 rounds of protein transductions only peaking at a maximal efficiency of 0.032%, inferred by AP staining. Moreover, the six rounds of transductions take place over an 8-week protocol, with primary colonies being picked approximately 56 days following the initial transduction (Kim et al., 2009). This low

efficiency, again, was said to be in relation to the proteins' continual endosomal entrapment. A number of studies have been carried out since these seminal pieces of work, yet only one publication has reported a reprogramming efficiency over 0.05% in humans (Lee et al., 2012). In theory, the process of reprogramming using purified proteins is ideal in terms of the progression of iPSCs to the clinic. Transient protein transduction is a significantly safer method of iPS cell production as opposed to exogenous pluripotency transgene expression. Its ease of manipulability permits the editing of pluripotency factors for transduction. Likewise, its protocol for iPSC production allows precise control of both protein dosage and timeframe of reprogramming factor expression. Yet, the power of complete control held within this method is outweighed by a number of bottlenecks. Protein reprogramming functions at a much lower efficiency than other widely accepted vectors and reprogramming methods. In addition, the proteins demonstrate manufacturing variability, poor in-vitro solubility/stability and perpetual endosomal entrapment, all of which are confounding factors to the vectors minimal uptake within the reprogramming field (Borgohain et al., 2019).

#### 1.2.2.3.4. **Transposon Vectors (PiggyBac & Sleeping beauty):**

Transposons are non-viral, DNA vectors which typically consist of a transposition DNA sequence of interest and an additional transposase expression cassette. Transposons, in comparison to other DNA vectors, have been demonstrated to possess a high transfection efficiency, a reduced immunogenic response and a versatility in terms of its genetic payload or cargo (~10kb). That said, the transposon system relies on the integration of its expression cassettes into the host cell genome to permit prolonged gene expression (Rao and Malik, 2012). However, the virtue with transposon vectors is that once the required expression period is complete, transient re-introduction of the transposase enzyme permits the excision of the integrating vector leaving the cells genome intact. On the whole, transposons provide a system which is inexpensive, easy to purify, demonstrates a high transfection efficiency and can stably express a gene of interest without leaving any permanent modifications - providing a good candidate for iPSC production and clinical progression.

As such, two main transposon vectors have been utilised in iPSC research, *PiggyBac* (PB) and *Sleeping beauty* (SB). Both of these transposon systems consist of a donor plasmid – containing the gene(s) of interest to be introduced into the host cell genome. These donor plasmids are flanked by inverted terminal repeats (ITRs) that are critical to its transposition. Finally, as mentioned, a helper transposase-expressing plasmid is also co-transfected alongside the donor plasmid. The transposase enzyme functions by catalysing the ITR sequence to isolate the gene/sequence of interest before integrating it into the genome of the host cell via a '*cut and paste'* mechanism (Haridhasapavalan et al., 2019). Few differences exist between the PB and SB transposon vectors. The PB transposon donor integrates in a semi-random fashion, being more inclined to integrate within transcription units in the host cell genome (Hu, 2014b). Meanwhile, the SB system integrates randomly within the hosts genome, displaying no propensity for specific genes/gene regulatory elements. However, both systems have, from a single transfection, been demonstrated to be able to develop, and ultimately be excised from, murine and human iPSCs (Kaji et al., 2009; Woltjen et al., 2009; Muenthaisong et al., 2012; Davis et al., 2013). This, however, is a far from perfect method of iPSC production. Firstly, both the PB and SB systems reprogram with low efficiencies – with the PB method being reported to reprogram at efficiencies of 0.02-0.05% and the SB system at ~0.02-0.03% (Haridhasapavalan et al., 2019). Moreover, the transposon system functions on the fundamental basis of integrating foreign DNA into the somatic cell genome. Ultimately, this may increase the risk of insertional mutagenesis and chromosomal rearrangements. Despite the fact that the DNA can be excised with the transposase enzyme at the end of the protocol, thorough screening of cells produced using this vector is still a necessity. In addition, reports also suggest that with both transposon systems, that the transposase excision reaction occurs with a 95% success rate, an important caveat, reiterating the necessity of screening protocols post iPSC production (Wang et al., 2008). As aforementioned, the extra-step adds further time and labour to the protocol, making it longer than procedures carried out for other non-integrating vectors (Igawa et al., 2014). That said, both of these systems have been demonstrated to produce integration-free iPSCs – an important merit not to be overlooked.

1.2.2.3.5. **Plasmid DNA:**

One of the most rudimentary methods to manipulate gene expression is with the use of plasmid DNA derived from bacteria. Plasmids have the capacity to replicate autonomously within the bacterium and can be both linear and circularised. However, the linear form is more susceptible to exonuclease digestion and so the circular plasmid is most often utilised (McLenachan et al., 2007). Plasmids have long been manipulated to over-express a gene of interest in a host cell and this method has been recognised as having clear potential for exploitation in the field of iPSC development. The Yamanaka lab was first to publish the successful derivation of iPSCs using transient plasmid transfections in murine fibroblasts (MEFs), with no signs of integration (Okita et al., 2008). The Yamanaka lab have recognised since their pioneering work first producing both mouse and human iPSCs, that for iPSC technology to have a clinical applicability, that a non-integrating vector must be utilised to minimise the risk of tumorgenicity. The lab utilised two plasmids, one expressing *Oct3/4*, *Sox2* and *Klf4* and the other plasmid expressing *c-myc*, all under a constitutively active *CAG* promoter. The plasmids likewise contained an ampicillin resistance gene. MEFs, modified to contain a *Nanog*-GFP reporter were subjected to multiple rounds of plasmid transfections. Four transfections were carried out in total, the OSK plasmid was transfected on days 1 and 3, while the *c-myc* plasmid on days 2 and 4. The plasmid was found to successfully produce iPSCs in 70% of reprogramming experiments with efficiencies ranging from 0.0001-0.0029% (Okita et al., 2008). It was clear from this initial experiment that the plasmid-based reprogramming system offered a much more clinically viable method of iPSC production, as opposed to integrating viral vectors. Yet, with multiple transfections required, increasing workload and cellular toxicity, and an extremely low reprogramming efficiency, it was clear that further work on plasmid-based iPSC production was required. Alterations to the stoichiometry of the plasmids utilised within the reaction were soon undertaken. Experiments were carried out utilising the key Yamanaka factors, each in separate plasmids, successfully producing human iPSCs, albeit again, at a considerably lower efficiency than other reprogramming methodologies (Si-Tayeb et al., 2010). It was recognised that polycistronic vectors, expressing multiple reprogramming factors could potentially increase the plasmid systems reprogramming efficiency. This would, in theory, increase the propensity of

transfected cells receiving the complete reprogramming transcription factor 'cocktail mix' and therefore increase the cells chances of undergoing complete pluripotency-recapture. Despite this, the use of polycistronic vectors was also viewed as a '*double-edged sword'*. The dogma central to successful reprogramming is having continual and balanced pluripotency transgene expression, sanctioning a cells progression through the three reprogramming phases. A polycistronic vector holds the potential to create an imbalance within the transgene stoichiometry which may result in non-beneficial effects on a cell's reprogramming potential. Moreover, having an all-encompassing reprogramming plasmid, would increase the size of the vector dramatically, which in-turn, would inversely decrease the transfection efficiency of the plasmid. This again would not hold any benefits in terms of increasing the reprogramming efficiency of the vector (Haridhasapavalan et al., 2019). However, the introduction of the picornaviral 2A system, brought about a '*happy medium*' in terms of altering the plasmid systems stoichiometry. The 2A system, in brief, permits the joining and simultaneous transcription of several transgenes producing a single transcript containing all said transgenes. The system is then however, self-cleaving, permitting no subsequent changes to the reprogramming factors eventual translation (Osborn et al., 2005). That said, the issues associated with stoichiometry are not the only sticking point in relation to the lack of success achieved using transfected plasmids. Plasmids are not self-replicating in eukaryotic cells, and as such the transient nature of the system is insufficient to induce a complete reversal of the differentiation process and onset of pluripotency. It is reported that a minimum of 12 days of continual pluripotency gene expression is necessary to reprogram and produce iPSCs, and that ultimately expression is likely also required for a period beyond this too, to ensure completion of the maturation-stabilisation phase transition (Brambrink et al., 2008; Si-Tayeb et al., 2010). Therefore, the transient nature of this vector is likely to owe to its reduced functionality. The presence of immunogenic plasmid backbone prokaryotic DNA sequences such as antibiotic resistance genes and unmethylated CpG dinucleotides will potentiate an upregulation of pro-inflammatory cytokines (including IFN-γ) from the host cells. This, in turn, will increase levels of plasmid removal and transgene silencing – contributing to the plasmids aforementioned transiency (Yew et al., 2000). It was clear that for plasmid to become

a more effective system within reprogramming, that changes to alter its transiency, and the aetiology of that transiency, were required.

1.2.2.3.6. **Minicircle:**

Being first described in the Crouzet lab, minicircle vectors were generated and found to transcend traditional expression plasmids by negating many issues associated with their use (Darquet et al., 1998). Conventional plasmids, through bacterial-associated DNA sequences, carry a risk of cytotoxicity, immunogenicity and uncontrolled therapeutic gene dissemination. Minicircles, however, are episomal, supercoiled circular DNA sequences which are devoid of almost all bacterial-associated DNA sequences including: the origin of replication sequence and antibiotic resistance genes (Hu, 2014b). Additionally, the minicircle system is reported to produce vectors which are smaller in size to traditional plasmids (~4kb in comparison to sizes as high as 15kb with plasmid) and seemingly have a greater transgene expression capacity both *in-vitro* and *in-vivo* too (Chen et al., 2003; Maucksch et al., 2009). Moreover, being smaller in size the vector demonstrates a greater transfection efficiency and likewise being devoid of pro-inflammatory bacterial sequences, possesses a much greater longevity of expression within its host cell in comparison to conventional plasmids (Darquet et al., 1998; Chen et al., 2003; Narsinh et al., 2011; Haridhasapavalan et al., 2019). It is clear from the above work, that the physical and functional properties associated with minicircle, make it a much more attractive candidate for the delivery of pluripotency transgenes in somatic cell reprogramming. Initially, successful reprogramming was carried out utilising a polycistronic, ‘reprogramming minicircle’ in adipose stem cells (Jia et al., 2010; Narsinh et al., 2011). The minicircle system however, still functioned at a low reprogramming efficiency, ~0.005%. This was disappointing, as not only did the vector fail to reprogram and produce iPSCs in an efficient manner, it did so on a cell type that is considered to have a greater plasticity and ease of reprogramming than fibroblasts (González et al., 2011). Moreover, upon reprogramming fibroblasts with the same vector, Jia et al, 2010, were able to produce *bona-fide* iPSCs but with a 10-fold drop in efficiency again. Additionally, the protocol undertaken in these experiments was laborious and potentially cytotoxic with a nucleofection of the minicircles on day 0, subsequent FACs sorting of successfully

transfected cells before a second and third transfection on days 4 and 6 following nucleofection (Jia et al., 2010). Despite this, the vector was able to produce transgene-free iPSCs which could express key pluripotency markers whilst also demonstrating an ability to also differentiate into cells of each of the three germ lineages. In the years and research efforts undertaken with minicircle since its first description in iPSC production, little improvement has been made in terms of improving its efficiency within the reprogramming process. This is despite changes being made to the delivery system, and its subsequent transgenes too. A report was published, however, using a 4 in 1 codon-optimized minicircle vector which was demonstrated to be able to reprogram and produce integration-free iPSCs in feeder-free and chemically defined conditions with a single transfection (Diecke et al., 2014). Duly, being able to reprogram and produce integration free iPSCs in this manner brings a high-level clinical relevance to this reprogramming method which should be considered seriously. Yet, with the efficiency of the process being so low, this is a big confounding factor which needs to be addressed in order to propel this method further into the clinical setting.

#### 1.2.2.3.7. **Episomal plasmid – *oriP*-EBNA1 system:**

As previously discussed, plasmid expression vectors are cheap, easy to produce, store and have a much longer shelf-life in comparison to other delivery vehicles utilised within the remit of reprogramming. Accordingly, *oriP*-EBNA1 episomal plasmids are an attractive candidate for the clinical production of iPSCs. Standard plasmid-based vectors are limited in their ability to reprogram somatic cells. This is often because the vector is highly transient, providing a transgene expression period that is insufficient to induce complete somatic cell reprogramming and subsequent iPSC production (Rao and Malik, 2012). To overcome this, episomal plasmids have been modified to contain both *oriP* (Origin of replication), and EBNA1 (Epstein-Barr nucleic antigen 1) and have since been demonstrated to reliably produce iPSCs from fibroblasts amongst a myriad of other cell types (Yu et al., 2009; Okita et al., 2013; Wang et al., 2013). The functional success of this system relates to the oriP-EBNA1 plasmid being capable of correcting transiency issues associated with standard plasmid-based vectors. Such corrections are conveyed by the modified vectors ability to tether to its host-cells somatic

chromosomes – a function conveyed by the Epstein-Barr virus (EBV)-derived EBNA1 protein. Moreover, the modified vector is able to replicate concurrently with somatic chromosomes during mitosis, again reducing any previous transiency-associated issues and thus permitting prolonged transgenic expression (Frappier, 2012). Modifications to the systems persistence and expression period have been critical to the success of the oriP-EBNA1 system.

Episomal tethering is, as mentioned, largely dependent on the viral EBNA1 protein which functions to temporarily unify chromosomal DNA and episomal plasmid DNA. The EBNA1 protein interacts at the bipartite locus *oriP* of the plasmid to maintain continual persistence and to also aid in plasmid replication within transfected cells (Hodin et al., 2013). The *oriP* locus consists of two main *cis* elements - as demonstrated in Figure 8 – the *Family of repeats* (FR) and *Dyad symmetry* (DS) (Lindner and Sugden, 2007). Both of these elements are, respectively, pivotal for the maintenance and replication of the plasmid and its subsequent transgene. As above-stated, the EBNA1 protein has binding properties, allowing it to interact with both the expression plasmid and the hosts chromosomal DNA. The binding of the EBNA1 protein to the expression plasmid is mediated by the FR region of the *oriP* locus. The FR, consisting of 20 tandem repeats of a 30bp sequence, contains around 20 high-affinity EBNA1 binding sites which mediates the binding of the viral protein to the plasmid (Hodin et al., 2013). The EBNA1 protein is likewise also capable of attachment to host cell genomic DNA via its protein N-terminal hook motifs, *LR1* and *LR2*, that bind to AT-rich somatic cell chromosomal regions (Hodin et al., 2013). Both of these interactions create a semi-permanent unification between the plasmid and host cell chromosomes (Sears et al., 2004). This interaction alone would permit greater vector persistence and subsequently, a prolonged transgenic expression. However, without an ability to replicate alongside the host cell, the plasmid, and its expression, would ultimately become diluted resulting in a far reduced reprogramming efficiency. Duly, the DS unit of the *oriP* bipartite locus, functions as an origin of plasmid replication and has been demonstrated to be essential to this function (Yates et al., 2000). Within its 65bp sequence, DS consists of 4 EBNA1 binding sites, alongside 3, 9bp sequences referred to as '*nonamers*' (Frappier, 2012). The binding affinity of these sequences for EBNA1 is relatively weak. Despite this, it has been suggested that both of these sequences are essential for DS function as a DNA replicator. There is, however, a

degree of ambiguity surrounding the exact function and necessity of the EBNA1 sequence within the initiation of this process, with conflicting reports being published. However, it is suggested that the EBNA1 protein is important in the recruitment of the origin recognition complex (ORC), which binds to the DS element of oriP, initiating host cell/plasmid DNA replication (Frappier, 2012). Moreover, reports have also demonstrated that the EBNA1 protein has direct interactions with proteins associated with ORC recruitment and DNA replication such as Cdc6 in murine models and Trf2/TRF2 likewise (Deng et al., 2002; Moriyama et al., 2012). It is therefore proposed that the DS unit of the *oriP* locus and the EBNA1 protein function with a dual-synergy and co-dependency in initiating the replication of the episomal plasmid within the host cell. Such co-dependency is further substantiated by reports demonstrating that EBNA1 bound to the FR region of *oriP* displays an inability to recruit ORC and initiate DNA replication – establishing the decisive role that the EBNA1-DS interaction has within this process (Schepers et al., 2001; Frappier, 2012). The confirmation of the DS locus is suggested to be the decisive factor to its function as a replicator – with the locus having a 21bp centre-to-centre spacing (Hodin et al., 2013). Duly, despite the exact mechanism remaining to be completely elucidated, it is experimentally clear that the addition of both oriP & EBNA1 to the expression plasmid system provide it with a prolonged capacity to divide and express its desired transgene.

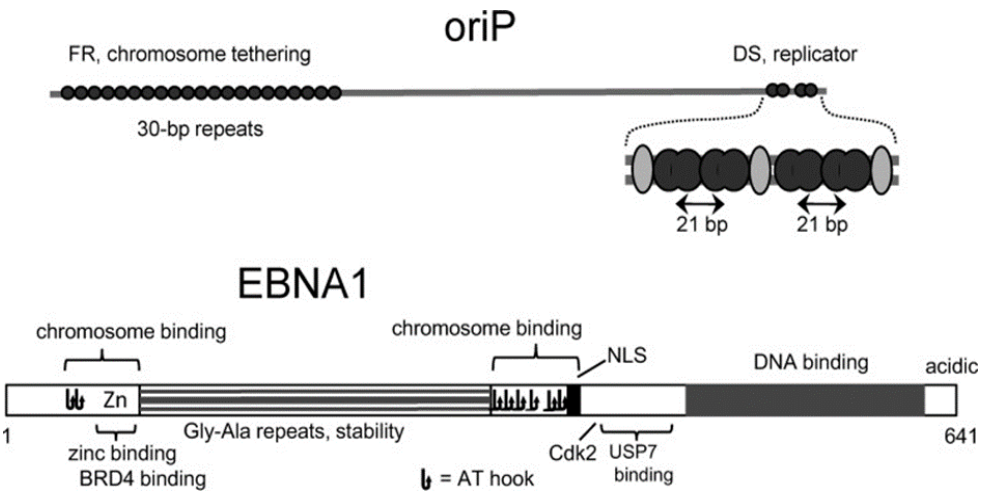



***Figure 8 - Features of both the oriP bipartite locus & the EBNA1 protein – source: Hodin et al, (2012).***

The attribute of having an increased transgene persistence but still being a non-integrating vector provides key functional properties to the oriP-EBNA1 system, that are desirable to the process of reprogramming and iPSC development. As such, the oriP-EBNA1 system has become a staple in the development of iPSC technology and has been successfully utilised within labs globally. The system currently relies on a single nucleofection of a 6-factor formula – OCT4, SOX2, KLF4, LIN28, L-myc & shp53. These six factors are split across three plasmid backbones. Nucleofection of an extra plasmid expressing the EBNA1 is also included to improve reprogramming efficiency by means of increasing the propensity of plasmid tethering and replication (Yu et al., 2009). The oriP-EBNA1 system utilises a CAG promoter to initiate and drive strong transgene expression. The CAG system is synthetically derived, utilising an early enhancer element from cytomegalovirus, the first exon and intron of the chicken β-actin gene alongside the splice acceptor of the rabbit β-globin gene (Wang et al., 2017). Each vector likewise also contains two of the six pluripotency factors separated by a 2A sequence. Derived from Picornavirus, the 2A sequence permits the co-expression of multiple genes (SOX2 & KLF4 in one vector (hSK) and L-myc & LIN28 (hUL) in another) by the well-characterised "*stop-go*" translational stuttering mechanism. The two transgenes, SOX2 & KLF4 for instance, are transcribed as a single unit, then, during translation the 2As "*self-cleaving*" properties come into fruition. At the ribosome, cleavage occurs between glycine and proline codons at the 2A C-Terminus upon translation (Wang et al., 2015). A proline codon directly after the 2A sequence permits the synthesis of two non-overlapping proteins. This is accomplished as the glycine codon promotes subsequent termination of translation before the "*stop-go*" system allows continued translation with proline permitting N-terminal formation of the next protein (Atkins et al., 2007). This subsequently allows the transcription and translation of two transgenes from a single promoter as depicted in Figure 9.

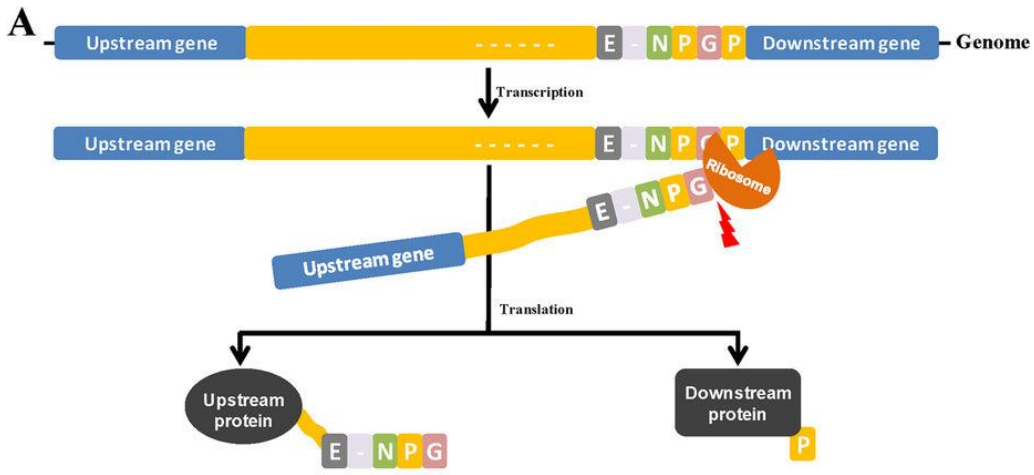



***Figure 9 - Mechanism of cleaving of the 2A system with an upstream and downstream gene to be expressed. The cleavage at glycine forms the C-Terminal of the upstream gene to be expressed. Whilst, the subsequent praline forms the N-Terminal of the downstream gene to be expressed (Source: Wang et al, (2015) online).***

The oriP-EBNA1 plasmid-based system has, duly, been uptaken in the reprogramming process as it displays a number of functional advantages which may favourably translate iPSC-technology into a clinical setting. The vectors prolonged expression alongside its reliability in producing iPSCs from a number of different cell types are clear incentives for the vectors proposed use. Moreover, the protocol outlined to employ the *oriP*-EBNA1 system into is extremely simplistic, and, although time-consuming, is relatively less labour intensive in comparison to some of the prior mentioned methods of iPSC production (Yu et al., 2009). Owing only to a single transfection, the *oriP*-EBNA1 operates with a relatively good efficiency as a non-integrating vector. The reprogramming efficiency of the system in xeno-free, chemically-defined conditions is however lower, with a report optimising this process finding that 9µg of vector being nucleofected into $1.0x10^5$ cells being optimal for colony formation. The lab was, with these conditions able to reach peak efficiencies of around 0.017% (Bang et al., 2018). However, this was only 14 days from the initial nucleofection as opposed to a standard protocol reaching up to 30 days. Despite this, it is possible to reprogram using chemically-defined, xeno-free products and gain iPSCs which are transgene and integration free on a regular basis (Yu et al., 2009; Lee et al., 2014; Schlaeger et al., 2015). This is obviously hugely beneficial in terms of progressing the highly anticipated but largely underused iPSC technology into the clinic and beyond.

Yet, the process of iPSC production with the *oriP*-EBNA1 system is not without potential pitfalls. Crucially, some of these issues associated with the vector could be limiting factors to the number of clinical grade lines produced in this manner. First, some of the transgenes expressed which permit the induction of pluripotency within somatic cells using the *oriP*-EBNA1 system can likewise infer the induction of gene networks associated with cancer development. MYC transcription networks contribute to the maintenance of iPSC self-renewal and iPSC multiplication, likewise, shp53 is reported to increase reprogramming efficiency, however both also harbour a potential for an increased tumorigenicity (Chin et al., 2009; Hong et al., 2009).

Secondly, the chromosomal binding nature of *oriP*-EBNA1 system which permits the production of iPSCs contributes, for the same reason, to the prolonged retention of the vector subsequent to iPSC production. Considerable ambiguity surrounds the theory and actuality of when iPSCs become plasmid deficient. Ultimately, prolonged retention of the vector reduces the clinical applicability of said iPSCs (Drozd et al., 2015; Schlaeger et al., 2015; Churko et al., 2017). Additionally, although rare, a chance exists that the *oriP*-EBNA1 system may integrate into the hosts somatic genome, increasing the work capacity of the reprogramming process as any iPSCs produced using this system must be screened for any vector-derived integrations (Churko et al., 2017; Haridhasapavalan et al., 2019). Despite the possibility of isolating transgene free iPS cells, the persistence of the plasmid limits which iPS cells can be used clinically, since cells with persisting vector and transgene expression or potentially containing integrations could have adverse downstream inflammatory/innate immune responses and tumorigenic implications (Yoshinda & Dowdy, 2017).

The EBNA1 protein, as previously mentioned, is critical to the improved functionality and applicability of expression plasmids within the realm of reprogramming and pluripotency induction within somatic cells. The system is robust and can produce iPSCs that meet cGMP requirements (Baghbaderani et al., 2015). Yet, the inclusion of the volatile viral protein, EBNA1, within the vector is both a *blessing and a curse*, as expression plasmid retention is dependent on the protein's functionality, yet it has been clearly identified to potentiate a myriad unwanted effects within transfected cells (Kennedy et al., 2003; Hong et al., 2006; Gruhne et al., 2009; Valentine et al., 2010; Pannone et al., 2014). As such, a fine balance exists with the use of this vector with

regards to honing its ability to produce iPSCs whilst also turning a *blind eye* to its unequivocally detrimental side effects.

Despite this, clinical trials that are currently ongoing and receiving approval using iPSCs have almost always been reprogrammed using the *oriP*-EBNA1 vector. Currently, as of mid-2019, there are nine clinical trials utilising iPSC-derived cell therapies (Vanneaux, 2019). The first was approved in 2014, using autologous iPSC-derived retinal pigment epithelium (RPE) cells which were transplanted as a sheet, with the aim to reconstitute sight in age-related macular degeneration (AMD) patients (Mandai et al., 2017). Since then, trials using iPSC-derived cell therapies for treating a number of disorders, from Parkinson's disease, spinal cord injury and thrombocytopenia, as well as producing iPSC-NK cells (FT500) for immunotherapeutic purposes, have arisen (Kikuchi et al., 2017; Akabayashi et al., 2018; Crow, 2019; Nagoshi et al., 2019). There are however, more than double the amount of trials utilising ESCs than iPSCs, which when considering the issues surrounding their derivation and increased susceptibility of rejection in comparison to iPSCs, can seem surprising. A number of issues have stunted the clinical delivery of iPSC-cell therapies such as tumorgenicity in relation to both transgenes and integrations, alongside the scalability of clinically relevant iPSCs to meet therapeutic demand (Vanneaux, 2019). It is clear that there are still a number of limiting factors with regards to iPSC exploitation that can be related to the choice of vector to reprogram and produce iPSC lines with.

Ultimately, it is clear that the choice of vector is critical, not just in determining the process and efficiency of iPSC development, but also their quality, which in turn has huge implications on the cells prospective use. The *oriP*-EBNA1 system is a robust, efficient and practical system in terms of iPSC development, but is by no means perfect. The ideal modality for generating iPSCs would be a robust, efficient method, based on a non-viral, non-integrating, transient vector. Such vector would have minimal augmenting elements, lack any pro-inflammatory additional sequences and non-specific modulatory effects. Indeed, the discussion offered thus far has demonstrated why the *oriP*-EBNA1 system has become the main preference amongst a plethora of other reprogramming systems. The vector has helped to progress iPSC-related therapies and is subsequently providing the foundations for potentially life-changing treatments. That said, the vector does have its downfalls and so, despite the

systems successes, these issues should not be ignored and if possible, be corrected for with a vector type that can match the *oriP*-EBNA1 systems functionality, whilst also negating some of the above-mentioned issues.

***Table 2 - Overview of reprogramming methods.***

| Vector | Advantages of use within stem cell research | Potential clinical impediments | References |
|---|---|---|---|
| Retrovirus (MMLV) | ○ Efficient and widespread infection in target cells.<br>○ Stability & longevity in expression. | ○ Genomic integration.<br>○ insertional mutagenesis.<br>○ Re-activation of silenced transgenes.<br>○ increased propensity tumorigenesis.<br>○ Viral proteins increase the propensity for inflammatory/innate immune response.<br>○ Transduction limited to dividing cells. | (Takahashi et al., 2007; Lee et al., 2012; Hu, 2014a) |
| Lentivirus (HIV) | ○ Efficient and stable transduction.<br>○ Dividing and non-dividing cells transduced.<br>○ Potential for integrase deficiency<br>○ limited genomic integration. | ○ Genomic integration.<br>○ Insertional mutagenesis.<br>○ Inflammatory response associated with viral transduction.<br>○ delayed iPSC development. | (Maherali and Hochedlinger, 2008) |
| Adenovirus | ○ Stable exogenous expression within target cells.<br>○ Low-level genomic integration. | ○ Altered reprogramming kinetics.<br>○ Highly inefficient in relation to other viral derived vectors.<br>○ Development of Tetraploid iPSC lines.<br>○ Possibilities for adenoviral DNA integration. | (Rao and Malik, 2012; Mora et al., 2017) |
| Sendai Virus | ○ Relatively high reprogramming efficiency.<br>○ Reliable colony production. | ○ The vector is slow in clearing – high passage cells still contain the viral vector.<br>○ Reduced ability to manipulate deliverable factors. | (Schlaeger et al., 2015; Churko et al., 2017) |

| | | | |
|---|---|---|---|
| | ○ Non-integrating RNA vector.<br>○ Single transduction required.<br>○ Little workload for iPS cell production. | ○ Presence of viral proteins undesirable for clinical translation.<br>○ Single clinical vendor thereby increasing expense.<br>○ Lack of clinical grade iPS lines. | |
| mRNA | ○ No genomic integration.<br>○ Vector transiency improves clinical applicability.<br>○ Shorter reprogramming duration.<br>○ Relatively efficient within colony production.<br>○ Low donor cell level required.<br>○ Relatively low aneuploidy rate. | ○ Increased workload.<br>○ mRNA extremely transient so multiple transfections required (up to 17 transfections).<br>○ Difficulties within reprogramming primary fibroblasts.<br>○ Difficulties in the reproducibility and implementation within laboratories.<br>○ No validation or reported iPSC production from blood. | (Hu, 2014a; Schlaeger et al., 2015; Mora et al., 2017) |
| miRNA | ○ Transient system<br>○ Demonstrated to reprogram fibroblasts to bona-fide iPSCs<br>○ Can completely replace the OSKM system | ○ Transient system which requires multiple transfections making the protocol laborious.<br>○ Reprograms with a low efficiency (~0.01%)<br>○ Can function with a higher efficiency but requires permeant integration into host cell genome. | (Miyoshi et al., 2011; Borgohain et al., 2019) |
| Recombinant proteins | ○ Extremely safe system of iPSC development<br>○ Transient.<br>○ Ease of manipulability for different reprogramming factors. | ○ Protein expression is transient – often being trapped in host cell endosomes reducing functionality.<br>○ Protocol requires multiple transfections and is also a long process – laborious for lab users. | (Zhou et al., 2009; Lee et al., 2014; Borgohain et al., 2019) |

| | | | |
|---|---|---|---|
| | ○ Protocol allows complete control of protein dosage and timeframe of dosage. | ○ Great variability in success levels between labs globally. | |
| Transposons | ○ Easily excisable system despite being integrative.<br>○ Single transfection reducing labour of protocol.<br>○ Can transport large cargo in excess of 10kb. | ○ Risk associated with re-integration of the system and subsequent downstream insertional mutagenesis-related issues.<br>○ Low reprogramming efficiency (~0.05%).<br>○ Time consuming iPSC analysis to ensure integrity is imperative.<br>○ Additional excision step required. | (Borgohain et al., 2019) |
| Plasmid | ○ Inexpensive<br>○ Easy to produce and manipulate<br>○ Easy to store | ○ Reprograms with an extremely low efficiency (~0.0029%).<br>○ Transient system requiring multiple transfections.<br>○ Pro-inflammatory bacterial sequences increase propensity of transgene silencing and plasmid removal.<br>○ DNA-based system carries rare chance of integration. | (McLenachan et al., 2007; Si-Tayeb et al., 2010; Haridhasapavalan et al., 2019) |
| Minicircle | ○ Non-integrating.<br>○ Minimal bacterial sequences within vector limiting transgene silencing.<br>○ Transient by nature minimising vector persistence. | ○ Very low efficiency in adipose stem cells (0.005%).<br>○ Ten times lower efficiency in neonatal fibroblasts.<br>○ Little disparity between Minicircle and conventional plasmid efficiencies.<br>○ No confirmed reprogramming in adult hDFs.<br>○ Minimised efficiency increases need for multiple transfections.<br>○ Vector is bacterially amplified increasing propensity for retention of bacterial epigenetic sequences. | (Hu, 2014a; Mora et al., 2017) |

| | | | |
|---|---|---|---|
| *oriP*-EBNA episomal plasmid | ○ No genomic integration.<br>○ Can be transient to produce “footprint free” iPSCs.<br>○ Easily produced and cheap.<br>○ Simplistic protocol and limited reprogramming workload promoting global utilisation. | ○ Inefficient (0.01% efficiency in fibroblasts).<br>○ Relatively high rates of aneuploidy (~11.5% of iPSCs produced).<br>○ Slow removal of vector – more than a 1/3 of iPSCs produced still contain reprogramming vector beyond P11.<br>○ Increased propensity for transgene silencing with bacterial sequences being present.<br>○ Persistence of bacterial sequences and viral proteins disadvantageous for clinical progression.<br>○ Chance of genomic integration, albeit rare. | (Schlaeger et al., 2015; Churko et al., 2017; Mora et al., 2017; Haridhasapavalan et al., 2019) |

## 1.3. **Doggybone (dbDNA) clinical grade vector:**

Produced by Touchlight Genetics and patented since 2008, Doggybone DNA (dbDNA) is a novel vector. The dbDNA system utilises clinical grade DNA, producing a double-stranded closed linear construct with a flexible transgenic cassette. dbDNA vectors are subsequently free from bacterial CpG islands/sequences as well as antibiotic resistance genes, thereby minimising any potentially immunostimulatory effects of the vector within its host cell. A premise therefore exists to test the dbDNA system within reprogramming and iPSC production and to note how it performs in comparison to other reprogramming methods to elucidate any benefits and disadvantages this new system may have.

### 1.3.1. **Manufacture:**

Doggybone DNA vectors are produced enzymatically via rolling circle amplification (RCA) (Figure 10) with no need for bacterial augmentation. The core method utilises the activity of two main enzymes; a Phi29 DNA polymerase and a protelomerase or telomere resolvase. Phi29 DNA polymerase is adopted within this process due to its

high fidelity (error rate of $1x10^6$ – $1x10^7$) as well as its high processivity (~70kbp). These features pertain to a high-suitability of the polymerase for the large-scale production of *good manufacturing practise* (GMP) DNA. The process of dbDNA manufacture, in brief, begins with a circular double-stranded DNA molecule (e.g., a plasmid) which is utilised as a starting template. The sequence of interest within the DNA template should be flanked by a 56bp palindromic protelomerase recognition sequence on either side. Once the starting template is denatured, the addition of the Phi29 DNA polymerase will then initiate RCA, resulting in double-stranded concatemeric repeats of the original template. A protelomerase/telomere resolvase is then added, the enzyme will subsequently bind to its recognition sites which flank the templates sequence of interest. The protelomerase after recognising inverted palindromic DNA sequences will catalyse strand breakage, strand exchange and ultimately DNA ligation. This results in multiple monomeric double-stranded, linear, covalently closed DNA constructs. The successful formation of dbDNA closed-ended structures makes the vector resistant to exonuclease activity. The DNA outside the gene of interest (e.g., the original vector backbone) will likewise be concomitantly processed by the protelomerase enzyme. However, these regions are removed by the sequential action of restriction enzymes cutting at restriction sites unique to the vector backbone before exonuclease digestion of the released fragments, leaving only the covalently closed linear DNA containing the target sequence intact. dbDNA is then purified from small fragments and reaction components using size separation to leave only the dbDNA sequence of interest. The resulting dbDNA™ constructs can be used as a starting material for further amplification reactions if required. Ultimately, the final dbDNA constructs are minimal, containing only the sequence of interest. They are also therefore free from bacterial *CpG* motifs and the requirement for antibiotic resistance genes necessary for bacterial amplification. The dbDNA product will then undergo a number of quality control (QC) checks to ensure the purity and quality of the product. The vector undergoes Sanger sequencing to ensure sequence homology. The DNA quality will be determined from running the vector on an agarose gel ensuring uniformity, before a western blot is carried out to check for the removal of any processing enzymes. Endotoxin testing is likewise carried out to ensure a negligible or null presence. The vector purity is then likewise determined using readings from a spectrophotometer. The manufacturing process of dbDNA is rapid, taking only up to 2

weeks to produce clinical grade, cGMP compliant DNA that has undergone rigorous quality control checks. Production is also easily scaled-up with the possibility of producing grams (g) of a final product. Crucially, the process of vector production and quality control is cost effective, providing no limitations for its research or clinical use on a global basis.

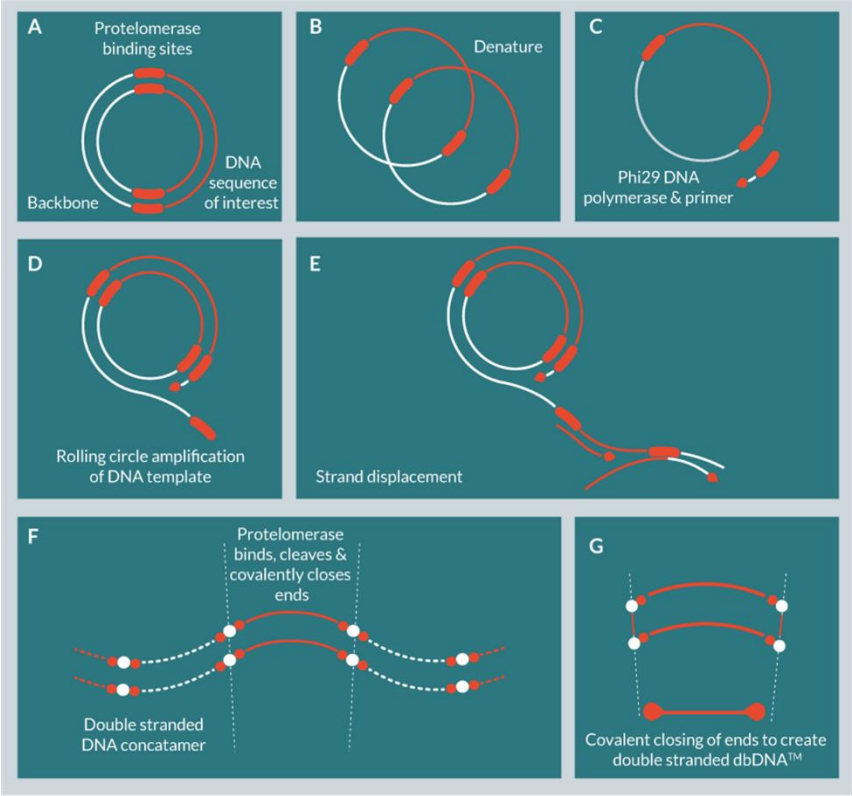



***Figure 10 - dbDNA manufacturing process from amplification and strand displacement to creating the closed end double-stranded product (Source: Karbowniczek et al, (2017) online)***

### 1.3.2. **dbDNA reprogramming systems:**

The dbDNA reprogramming system utilises three separate vector constructs, similar in principle to the *oriP*-EBNA1 system. The *oriP*-EBNA1 approach permits the expression of 6 transgenes across 3 different vectors - hSK (SOX2 & KLF4), hUL (L-myc & LIN28) and hOCT4-shp53 (OCT4 & Short hairpin for p53). The dbDNA constructs utilise the same number of vectors, yet with the expression of only 5

transgenes, as the shp53 is not present within this system. The Doggybone system, alike to the *oriP*-EBNA1 system utilises three individual vectors (Figure 11) as oppose to one large polycistronic construct. (Hornstein et al, 2016).

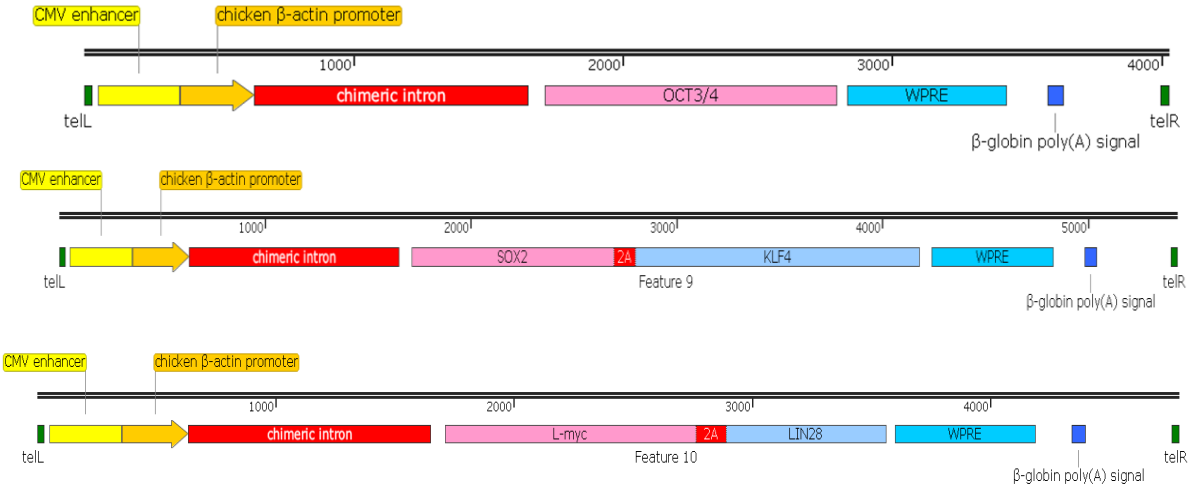



***Figure 11 - dbDNA vector contents for OCT3/4-shp53, hSK & hUL respectively.***

All of this is to say that, the use of dbDNA – insofar as it offers the possibility of a clinical grade, non-bacterial vector to deliver and initiate the de-differentiation of somatic cells – could herald the start of a newer, safer era for vectorology within reprogramming.

# 2.0. Results:

## 2.1. **dbDNA vector functionality, reprogramming potential and transcriptomic analysis:**

Previous literature provides a key insight into the importance of vector longevity and persistence of transgene expression in the process of cellular reprogramming. It is apparent that in order to achieve the re-constitution of an undifferentiated phenotype from a differentiated cell, that prolonged expression of pluripotency-related genes is critical in initiating and maintaining this process. Duly, from this point onward, the cell should then endogenously maintain pluripotent gene expression independent of transgenic expression. Any shortcomings within this process, in relation to the timeframe of transgene expression, will ultimately inhibit the formation of a true pluripotent phenotype, most likely resulting in an *in-between* or partially reprogrammed phenotype. As previously mentioned, modifications to standard expression plasmids had to be undertaken, adding *oriP* & EBNA1. This was in order to improve the longevity of the plasmid vector within host cells, to persist and express its pluripotency transgenes for long-enough periods to induce the reprogramming of differentiated cells back to pluripotency. Therefore, it is apparent that the timeframe of transgene expression by a vector is critical, not just to its success within the reprogramming niche, but to its eventual downstream translation progression likewise. Therefore, the analysis of vector kinetics, in relation to transgenic expression intensity and longevity can be insightful into vector functionality and its propensity for success within the reprogramming process. Moreover, the intensity of transgene expression, being able to significantly upregulate pluripotency gene transcription and ultimately protein expression, is likewise critical to the over-haul of a somatic cell's genetic '*machinery*', permitting the onset of pluripotency. It is important that any vector kinetics analysis therefore details the longevity of transgene expression, but also the intensity of such expression too.

Yet, the true value of the novel dbDNA system will be when applied to the process of reprogramming and incorporated into the publicised McKay lab protocol (Hawkins et al., 2016). The efficacy and efficiency of the vector to reprogram and produce iPSCs

will be determined within such experiments. Likewise, any cells produced should function to perpetually maintain an undifferentiated pluripotent phenotype but, also maintain an ability to differentiate into cells of the 3 germ layers following specific cues. The functional potential of iPSCs lies largely within the clinic, therefore, being able to produce iPSCs from differentiated somatic cells in *xenofree* conditions is likewise pivotal to the translational potential of this novel system too.

A transcriptomic microarray analysis would provide invaluable insights into global gene expression, permitting an understanding of more pronounced transitions, as well as more subtle changes to a cell's transcriptome. Alongside the development of the dbDNA vector to produce iPS cells, another core aim of the project was to determine any changes that the reprogramming delivery vector may have on its host cells internal environment. This was in relation to any benefit being incurred by the cell when reprogramming using the bacterial DNA-free clinical grade dbDNA system as opposed to the *oriP*-EBNA1 system which incorporated both bacterial and viral DNA sequences. Therefore, an evaluation of this type and scale would provide an extensive insight into key differences between these two vector types. Moreover, such analysis will help to guide further experiments to provide more substantial and supporting evidence for any identifiable changes.

Aims:

- Determine the vectors degradation kinetics in relation to transgene expression longevity and intensity for both the dbDNA and a plasmid system.
- Validate dbDNA reprogramming vectors structurally using restriction digests and functionally by determining the vectors ability to upregulate pluripotency proteins.
- Successfully carry-out the standard McKay lab reprogramming protocol using *oriP*-EBNA1 system.
- Carry out concurrent reprogramming experiments using both the dbDNA & *oriP*-EBNA1 expression systems on more than 3 different primary human dermal fibroblast cultures.

- Characterize any iPSCs produced for the expression of key endogenous pluripotency marker expression alongside the cells ability to differentiate into cells of all 3 germ lineages.
- Carry out reprogramming using an animal product or *Xenofree* protocol to determine the dbDNA systems robustness in terms of changing the conditions of the reprogramming process and denoting how this effects vector functionality in relation to iPSC development.
- Carry out reprogramming using the dbDNA system in a number of different somatic cell sources (Peripheral blood, urine-derived cells). Perform a microarray on iPSCs produced using the dbDNA & *oriP*-EBNA1 systems. Moreover, the inclusion of ESCs as a pluripotent gold standard control alongside the fibroblasts from which such iPSCs were derived was likewise important.
- Analyse the normalised microarray data to determine any global transcriptomic changes between both iPSC-types (dbDNA/*oriP*-EBNA1). This will help to deduce differences within the host cells reception of both reprogramming vectors.
- Begin to substantiate a mechanism behind any changes denoted between dbDNA and *oriP*-EBNA1 produced iPSCs determined from the microarray results.

### 2.1.1. **Comparison of vector degradation kinetics by quantifying transgene expression:**

To examine dbDNA transgene expression intensity and the longevity of its expression, we aimed to transfect a constitutively active GFP-expressing dbDNA vector into cells and quantify its expression level over a time-period. This was initially carried out in a Human embryonic kidney cell line, *HEK293T*, using a polyethylenimine (PEI) transfection-based system. In order to maximise the level of successful DNA transfection, a titration of different ratios of PEI to DNA (w/w) was undertaken. 1µg of DNA was transfected alongside varying ratios of PEI to determine the most optimal experimental conditions. GFP quantification was carried out to determine the lowest level of PEI accountable for the highest transfection and GFP expression levels, thereby limiting PEIs toxicity.

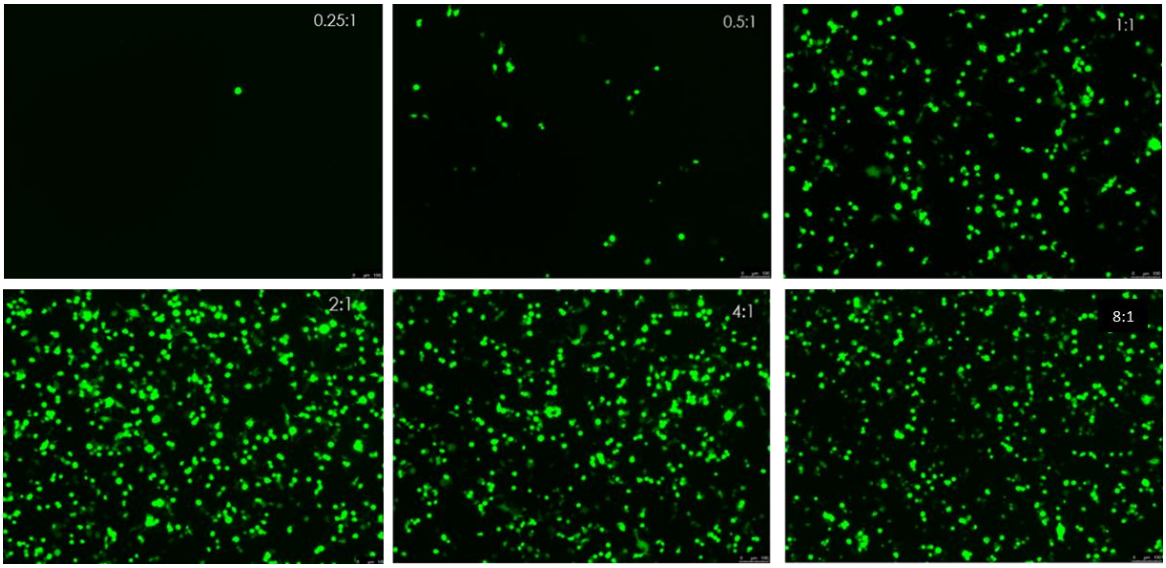



***Figure 12 - Fluorescent microscopy of GFP expressing cells at different PEI ratios.***

HEK293T cells were transfected with a consistent 1µg of dbDNA vector with different ratios of PEI. PEI:DNA ratios graduating from 0.25:1 up to 8:1 were incorporated and analyzed. Representative images of each PEI:DNA ratio (w/w) were taken at using a fluorescent microscope 24 hours following transfection. Scale bar represents 100µm. Experiment represents a single experiment.

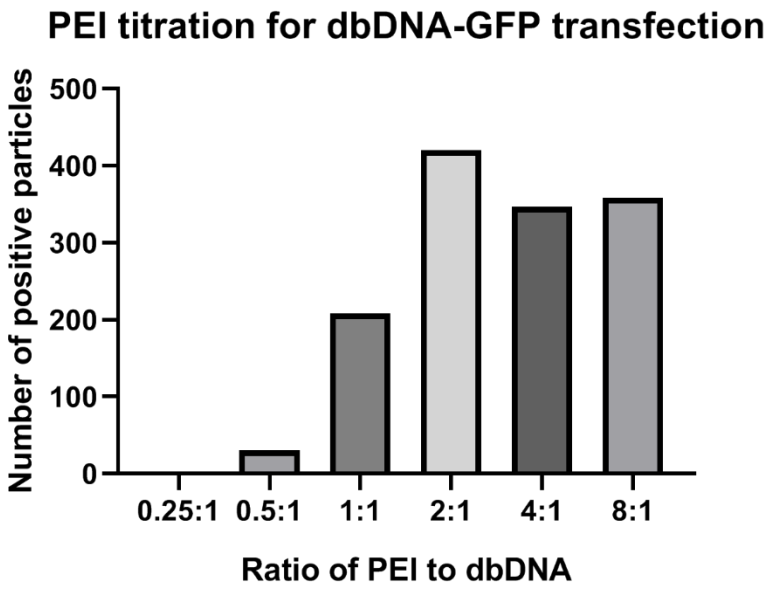



***Figure 13 – Quantification of positive particles per representative image using ImageJ.***

The representative images from the fluorescent microscope were analyzed using ImageJ to determine the number of positive particles (GFP expressing cells) per image. The number of positive particles was proportionate to the transfection efficiency. The analysis was carried out for each PEI ratio. The data represents a single experiment.

*Figure 12 & Figure 13* demonstrated on a qualitative and quantitative level respectively that a ratio of 2:1 (PEI: DNA (w/w)) resulted in the highest level of GFP expression. Following this, I then decided to then titrate the concentration (µg) of dbDNA, whilst maintaining the PEI ratio at a constant. This was carried out in *HEK293T* cells with care being taken again to isolate a concentration that provided maximal DNA uptake and GFP expression whilst again limiting cellular toxicity. The GFP expression within each DNA concentration was qualified and quantified in order to determine the most optimal conditions.

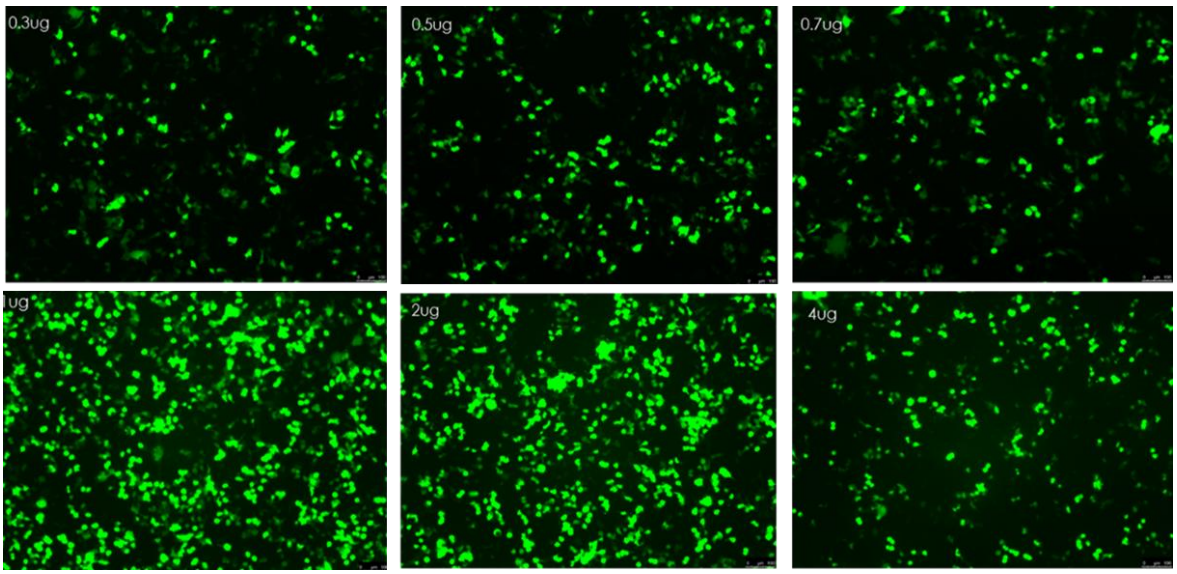



***Figure 14 – Fluorescent microscopy imaging of cells transfected with differing dbDNA concentrations.***

Concentrations of DNA from 0.3µg to 4µg were transfected into HEK293T cells in a 2:1 PEI to DNA ratio (w/w). Images of each DNA concentration were taken using a fluorescent microscope 24 hours post-transfection. Scale bar represents 100µm. Images represent a single experiment.

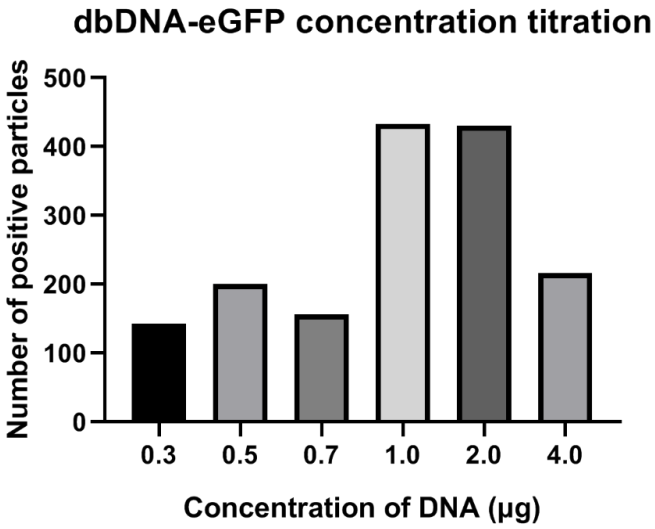



***Figure 15 - Quantification of positive particles per representative image using imageJ.***

The representative images from the fluorescent microscope were analyzed using imageJ to determine the number of positive particles (GFP expressing cells) per image. The number of positive particles was proportionate to the transfection efficiency. This analysis was carried out for each DNA concentration. Graphical representation of a single experiment.

Both 1µg and 2µg of DNA resulted in very similar levels of GFP expression within our predetermined PEI ratio (*Figure 14 & Figure 15*). However, 1µg of DNA was preferred, as it limited the total concentration of DNA being transfected; as DNA can likewise pose a threat of toxicity within host cells too. We clearly evidenced that a 2:1 PEI to DNA ratio (w/w) could be applied with 1µg of dbDNA for the most optimal vector transfection in *HEK293T* cells.

These initial experiments provided foundations which would permit further analysis into the vectors changes in longevity within the same cell type over time. An initial experiment was carried out incorporating molar eqivalent quantites of both the dbDNA and plasmid vector expressing a GFP transgene. The dbDNA system had a significantly reduced transfection efficiency in comparison to the plasmid system. Subsequently, net weight equivalent quantities were instead transfected, meaning that as the dbDNA system was often smaller than most plasmids, that more copies of dbDNA were transfected per cell. Only adopting this method were we able to acquire similar levels of transfection efficiency to the plasmid vector. It was decided that concurrent experiments employing the transfection of 1µg of both an eGFP-dbDNA vector and an eGFP-plasmid into *HEK293T* cells was to be carried out. The dbDNA vector being ~2.6kb in size and the plasmid ~4.7kb in size meant that according to Avogadro's number that there would be 1.79 copies of the dbDNA system per 1 plasmid copy ($1.53x10^{35}$ total copies of dbDNA per 1µg and $8.5x10^{34}$ total copies of plasmid per 1µg). Subsequently, the transfected cells were then analysed using flow cytometry to quantify the transfection efficiency of both vectors 24 hours post-transfection (Day 1). The degradation of expression could then be monitored, again, using flow cytometry to note the decline in GFP expression over time.

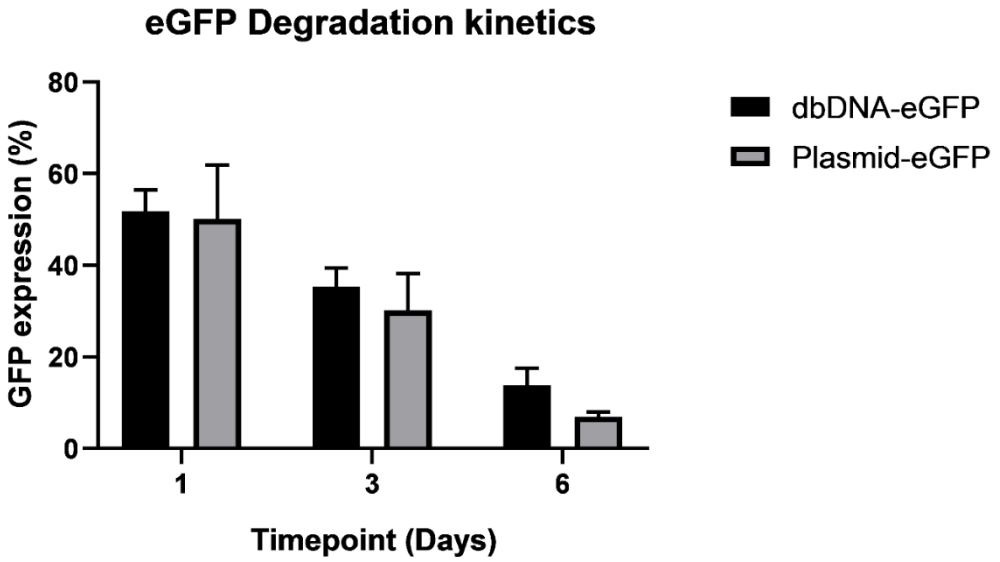



***Figure 16 -Analysis of GFP-expressing cells using flow cytometry.***

HEK293T cells were transfected with 1µg dbDNA-eGFP and 1µg eGFP expressing plasmid using PEI at a PEI:DNA ratio of 2:1. The cells were then analyzed for GFP expression using flow cytometry. This provided quantifiable data on the proportion of GFP expressing cells in a population. This was carried out 1, 3 and 6 days after the initial transfection. The data represents the transfection of a single well with error bars representing the SEM of the technical replicates.

The result demonstrated, over the 6-day analysis, that the dbDNA system was able to match the transfection efficiency of the plasmid-based system, and that its GFP expression declined at a slower rate in comparison to the plasmid vector too *(Figure 16*). This was an exciting result which could have profound implications with regards to the dbDNA systems downstream functionality within reprogramming and iPSC production.

In order to add more relevance to this first result, the same experimental conditions were applied over a longer time-period and to human dermal fibroblasts (hDFs) instead of *HEK293T* cells. Fibroblasts are the cell type most commonly utilised as the starting material for iPSC procurement. In order to mimic the reprogramming procedure carried out in the McKay lab, a total of 8µg of both eGFP-dbDNA & eGFP-plasmid was introduced into fibroblasts by *Amaxa* nucleofection. The fibroblasts were then analysed using flow cytometry to determine the transfection efficiency and monitor the longevity of GFP expression. As well as this, the intensity of such GFP expression was also determined by quantifying the *Median Fluorescence Intensity*

(MFI) of each GFP-expressing cell too. The same vectors were utilised as per the prior experiment, meaning that again there would be around 1.79 copies of dbDNA per 1 plasmid copy ($1.21x10^{36}$ copies of dbDNA in 8µg and $6.77x10^{35}$ copies of plasmid per 8µg).

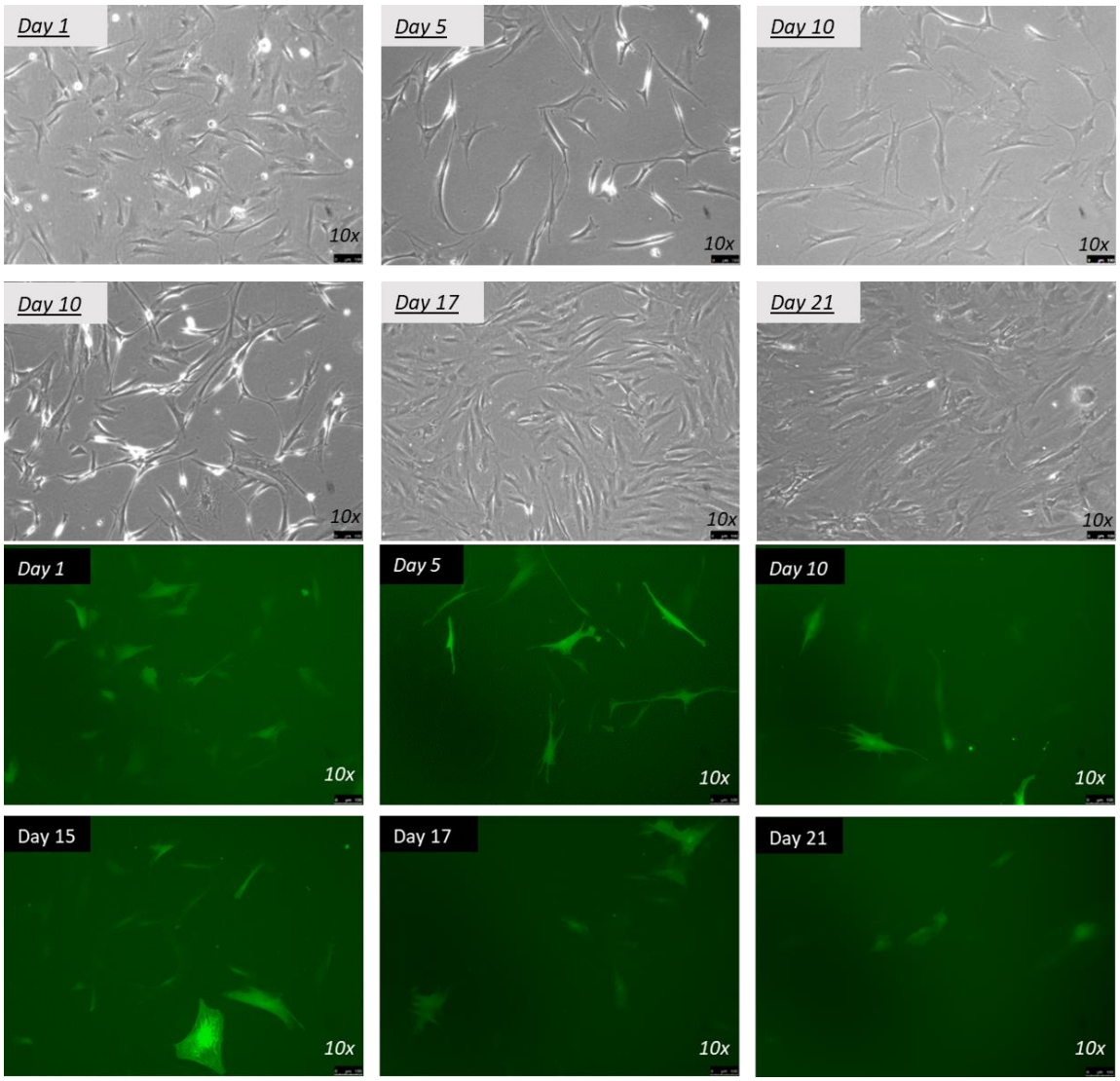



***Figure 17 - Representative images of eGFP-dbDNA transfected hDFs.***

Images represent the nucleofection of 8µg of dbDNA-eGFP into human dermal fibroblast cells. The images depict both phase and fluorescence for the dbDNA transfected hDFs from Day 1 to Day 21 post initial nucleofection. Scale bar represents 100um.

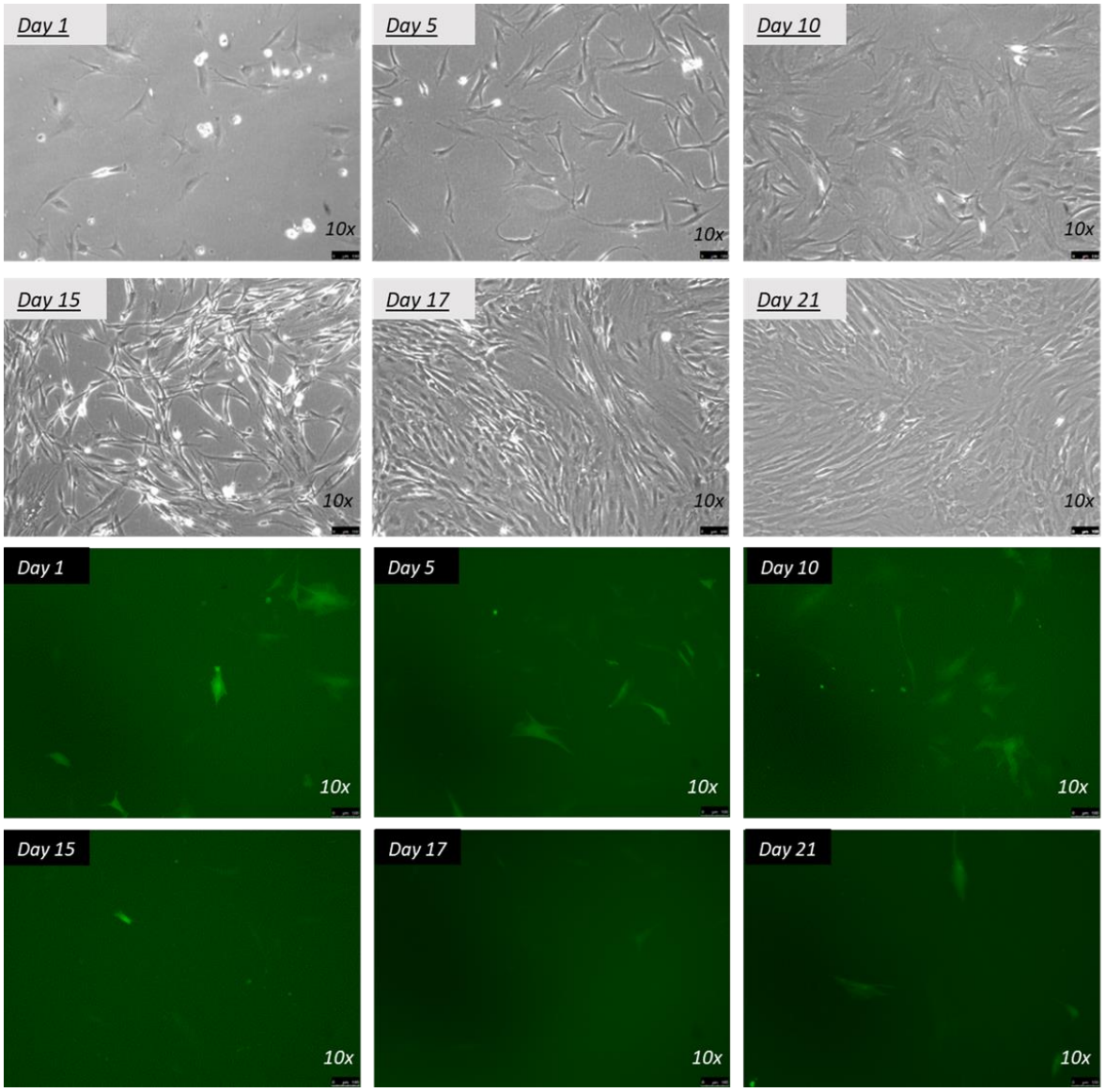



***Figure 18 - Representative images of eGFP-plasmid transfected hDFs.***

Images represent the nucleofection of 8µg of plasmid-eGFP into human dermal fibroblast cells. The images depict both phase and fluorescence for the plasmid transfected hDFs from Day 1 to Day 21 post the initial nucleofection. Scale bar represents 100um.

Representative images from this experiment provided qualitative evidence demonstrating a more persistent GFP expression in dbDNA transfected fibroblasts over time *(Figure 17 & Figure 18*). Such images helped to corroborate further data achieved when analysing the same cells, at the same timepoints, using flow cytometry to quantify GFP expression and MFI.

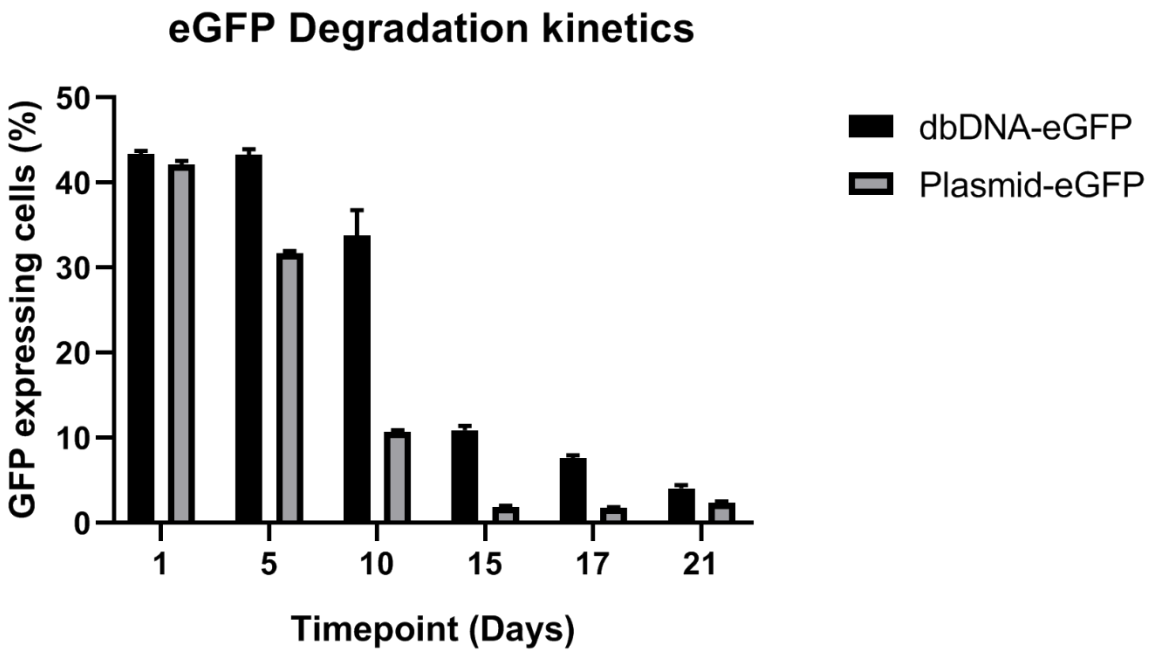



***Figure 19 -A flow cytometry analysis of GFP expressing populations following nucleofection with either dbDNA-eGFP or Plasmid-eGFP.***

8µg of both dbDNA and plasmid GFP-expressing vectors were nucleofected into human dermal fibroblast cells. The cells were taken for analysis over a time-period from 1 day to 21 days following initial nucleofection. The cells were analyzed using flow cytometry to quantify the proportion of GFP expressing cells within a complete population for each vector type. This population was then represented as a % proportion of the entire cell population. The data represents an n=1 with 3 technical repeats of the biological replicate. Error bars represent the SEM for the technical replicates.

On day 1 post-transfection, it was clear that both the dbDNA and plasmid system nucleofected with equivalent efficiencies. The subsequent timepoints demonstrated that the dbDNA system maintained its GFP expression at a greater rate and for longer a period of time. This was most apparent at Day 10 when the dbDNA system had a 3-fold higher GFP expression in comparison to its plasmid counterpart (*Figure 19*). This dataset substantiated the previously found data in *HEK293T*s and corroborated the representative images taken prior to flow cytometry analysis. It was clear that the dbDNA system can persist and express its transgene for longer time periods in comparison to standard plasmid systems. However, the intensity of transgenic expression was still to be elucidated between the two vector systems.

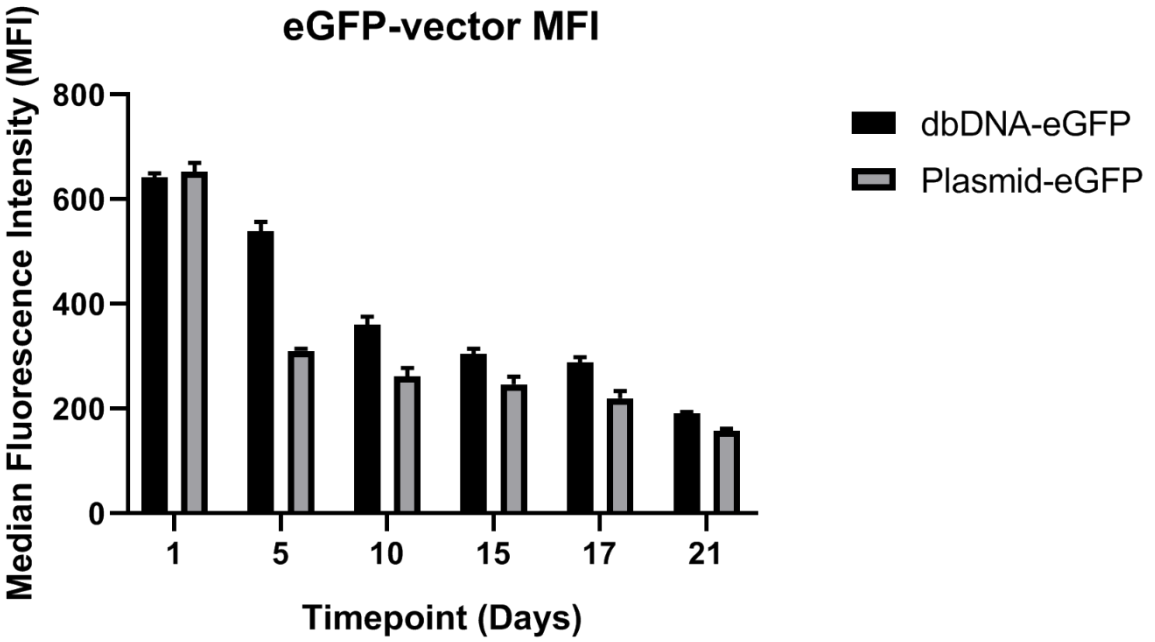


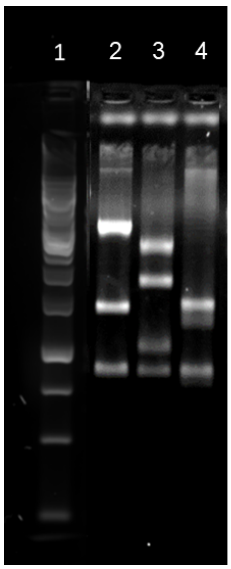

***Figure 20 - MFI of GFP expression from both dbDNA & plasmid vectors following flow cytometry analysis.***

Fibroblasts transfected with 8µg of either dbDNA or plasmid were analyzed using flow cytometry. Once the GFP expressing population of cells was determined, the median fluorescence intensity (MFI) could then be quantified and was proportionate to the brightness of such GFP expression. The data represents an n=1 with 3 technical repeats of the biological replicate. Error bars represent the SEM for the technical replicates

As with GFP degradation, the MFI expression demonstrated a similar trend, with GFP being expressed at a greater intensity and for a longer timeframe in dbDNA transfected cells as opposed to the plasmid system (*Figure 20*). Taken in its entirety, the experiment evidenced that the dbDNA system nucleofected with similar efficiencies, maintain its transgenic expression for a longer time period and expressed its transgene with a greater intensity over a longer timeframe too. Such facets are key to any downstream use of the dbDNA system within reprogramming and iPSC development.

### 2.1.2. **Functionality and structure of dbDNA reprogramming vectors:**

The results pertaining to the degradation kinetics study provided clear experimental evidence that the dbDNA system expressed its transgenic cassette with a greater longevity in comparison to a traditional plasmid. This was an exciting result which demonstrated key aspects of the dbDNA systems functionality which would favourably benefit its potential application to the process of reprogramming and iPSC development. Subsequently, Touchlight Genetics worked arduously, manipulated their methodology of dbDNA production and incorporated key pluripotency transcripts necessary for the reprogramming of somatic cells. This resulted in the production of 3 novel vectors for the purpose of this project (*Figure 11)*. The dbDNA vectors contained transgenes for key pluripotency markers: *OCT4, SOX2, KLF4, LIN28 & L-MYC*. Importantly, however, the dbDNA system omitted the inclusion of the *shp53* transgene that was expressed in the *oriP*-EBNA1 vector. Moreover, the dbDNA system also contained no bacterial DNA backbone which was likewise present in the *oriP*-EBNA1 system. This left us with 3 dbDNA vectors, all of which expressed the dbDNA telomeric cap sequence (necessary to make the vector a closed system) as well as our pluripotency transgene of interest. The stoichiometry of each vector was the same as the EBNA1 system; the dbDNA vectors expressed the same pluripotent transgenes on each cassette (*SOX2 & KLF4 / LIN28 & L-MYC*) besides from the OCT4 vector which was null of the *shp53* sequence. Upon receiving the novel dbDNA reprogramming vectors, our initial experiments focused around making sure the vectors were structurally sound. A restriction digest was carried out using the *ApaI* enzyme and determined if the number of digests/DNA bands was as expected from the whole vector sequence provided by Touchlight Genetics.




| Vector | No of cuts | Fragment sizes |
|---|---|---|
| (1) O'generuler DNA ladder | n/a | n/a |
| (2) dbDNA - hUL | 3 | 2626, 1325, 836 |
| (3) dbDNA - hSK | 4 | 2035, 1642, 916, 836 |
| (4) dbDNA - OCT3/4 | 4 | 1388, 1109, 894, 720 |

***Figure 21 - Restriction digest of the three reprogramming dbDNA vectors using the ApaI restriction enzyme.***

Image depicts an agarose gel following the digestion of dbDNA reprogramming vectors with the ApaI restriction enzyme. Lane 1 depicts a DNA ladder, Lane 2 dbDNA-hUL vector, Lane 3 dbDNA-hSK vector and Lane 4 dbDNA-OCT3/4. The table in the image demonstrates the number of fragments and their expected sizes once digested with ApaI.

The digest evidenced that the vectors were structurally as expected and digest providing the right number of bands from the sequence of each vector provided by Touchlight Genetics (*Figure 21*). This provided us with confidence that the dbDNA vectors we received were structurally intact, however, before incorporating them into the reprogramming process it was pivotal that we first tested the vectors functionality. Each dbDNA reprogramming vector was transiently transfected into HEK293T cells using PEI. The cells were then harvested for protein before a western blot was undertaken probing for key pluripotency markers that spanned all 3 vectors: *OCT4, SOX2* & *LIN28*. Both *SOX2* and *LIN28* were chosen over *KLF4* and *L-MYC* respectively due to the basal constitutive expression of both genes within *HEK293T* cells.

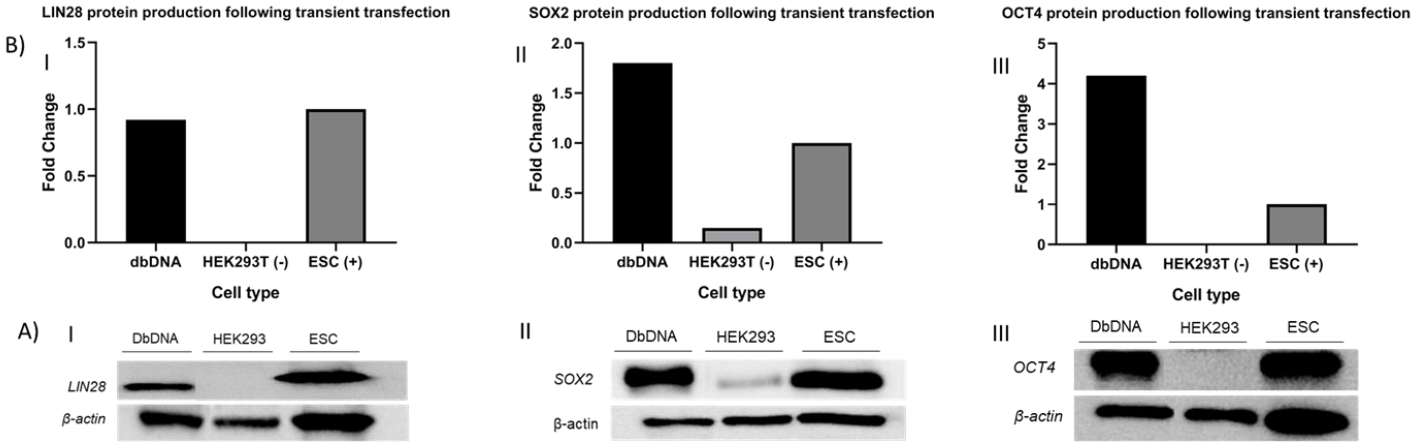



***Figure 22- Depiction of western blotting results from protein isolated from dbDNA transfected HEK293T cells for key pluripotency markers alongside their respective densitometry analysis.***

The result depicts the blot and bands for each cell type (dbDNA, un-transfected HEK293T & ESC) and for each pluripotency gene probed (‘*A)*’) alongside the respective densitometry analysis to quantify the level of protein upregulation (‘*B)*’). From ‘*I*’ to ‘*III*’ to right the genes probed include LIN28, SOX2 and OCT4 all of which are expressed from different vectors and transgene cassettes. Result depicts results from a single transfection and subsequent western blot.

The resulting blots demonstrated that the newly-produced dbDNA vectors were capable of upregulating the expression of key pluripotency genes to the level of that of ESCs; a pluripotent gold standard (*Figure 22*). This reassured us that the vectors we would begin working with were structurally and functionally apt and as such could begin to be incorporated into iPSC reprogramming protocols for functional analysis.

### 2.1.3. **Concurrent IPSC reprogramming experiments & vector efficiency:**

Having demonstrated that the dbDNA system was ready to be incorporated into the McKay Lab protocol for iPSC development, it was important to therefore be proficient with this process and protocol. Initial experiments incorporated 8µg total of the *oriP*-EBNA1 plasmid system into $4.5x10^5$ primary human fibroblast cells using nucleofection.

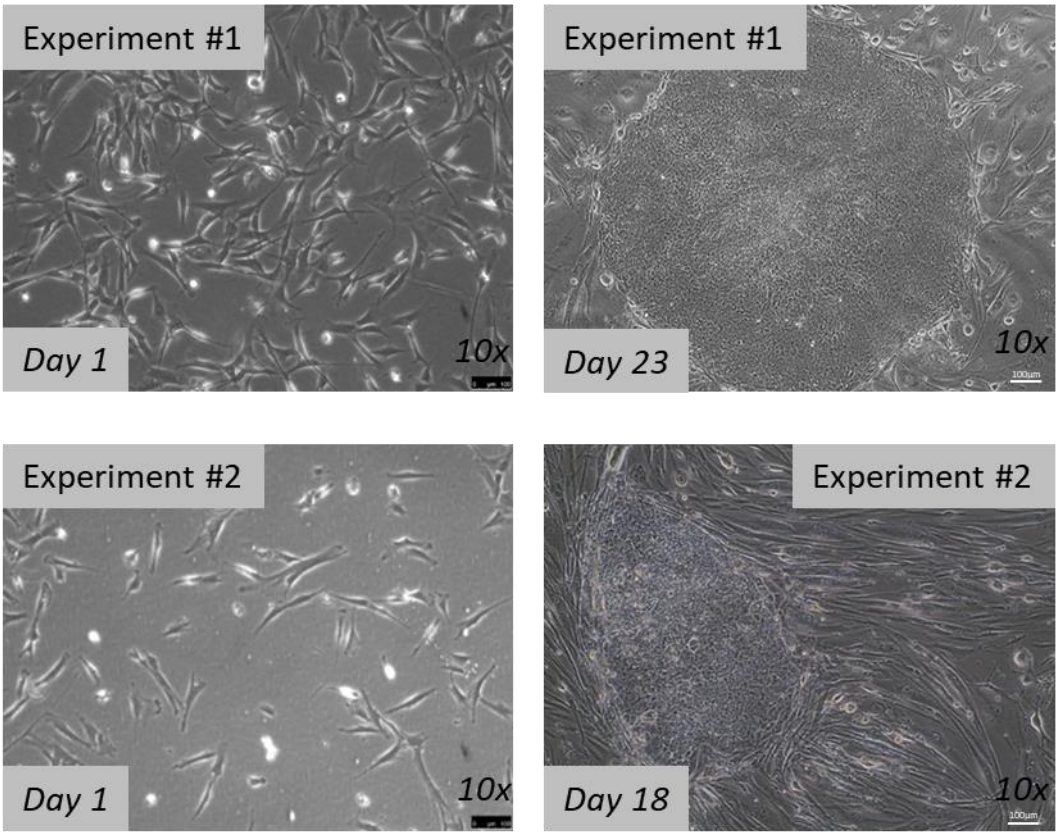



***Figure 23 - Successful preliminary reprogramming experiments carried out using the oriP-EBNA1 system.***

Images depict 2 different reprogramming experiments of 2 different primary fibroblast cultures using the oriP-EBNA1 system. The images detail the cells and their respective survival Day 1 post Amaxa nucleofection of the oriP-EBNA1 pluripotency vectors and subsequently upon primary colony development too.

The McKay Lab protocol was reproducibly successful when employing the *oriP*-EBNA1 system, resulting in the successful reprogramming of 2 different primary fibroblast cultures *(Figure 23*). Therefore, having successfully carried out the iPSC reprogramming process using the '*gold standard*' *oriP*-EBNA1 plasmid system and become accustomed with the protocol alongside the cellular changes associated with the process, I then wanted to apply the dbDNA reprogramming vectors to the same reprogramming protocol and conditions. The oriP-EBNA1 system incorporated the use of 4 different vectors which were 10,180bp (hSK/hUL/hOCT-shp53) and 5078bp (EBNA1) in size. A total of 2µg of each reprogramming vector was used in each reprogramming experiment (8µg sum total). Therefore, a total of ~$3.91x10^{35}$ copies of the oriP-EBNA1 plasmid was used in each reprogramming experiment. The dbDNA

system incorporated the use of 3 different vectors: dbDNA-hSK (5515bp), dbDNA-hUL (4873bp), dbDNA-OCT4 (4110bp). A total of 2333ng of each vector was applied to a reprogramming experiment (8µg sum total) meaning a total of ~5.84x10$^{35}$ copies of dbDNA reprogramming vectors were used in the same protocol. Therefore, there were 1.49 copies of the dbDNA system per 1 copy of the oriP-EBNA1 plasmid when carrying out their respective experiments.

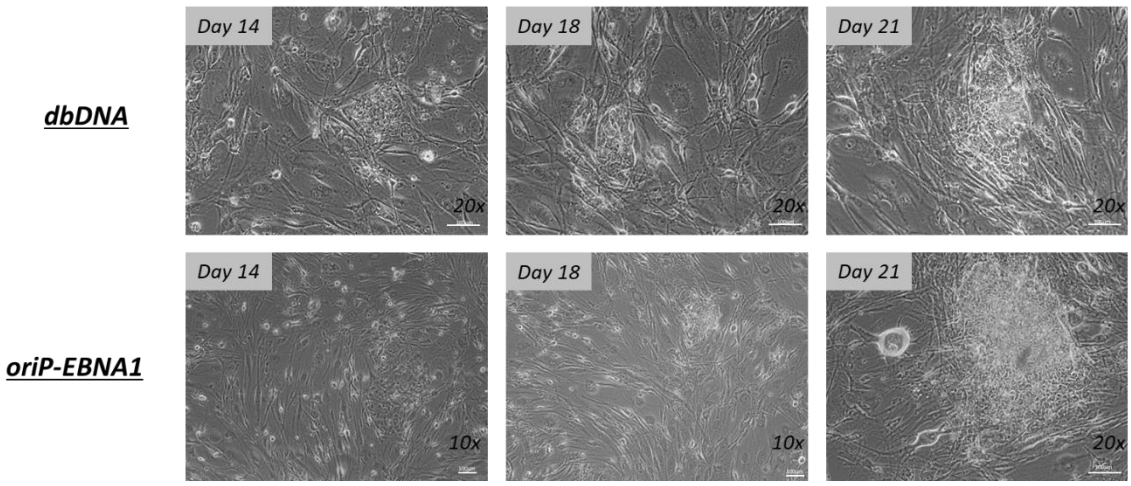



***Figure 24 - Reprogramming timeline of CLN3-hDFs using both dbDNA and oriP-EBNA1 vector systems until primary iPSC colony formation.***

Concurrent reprogramming of the same *CLN3* mutated primary fibroblasts using both the dbDNA and oriP-EBNA1 reprogramming vectors. Images depict a single experiment across several timepoints; from day 14 to day 21. Both vector types yielded primary colonies from the fibroblasts.

The first reprogramming experiment yielded what phenotypically resembled primary iPSC colonies in both the oriP-EBNA1 and dbDNA nucleofected fibroblasts (*Figure 24*). Primary iPSC colonies differ phenotypically from the surrounding fibroblasts, often growing in a compact circular organization with the cells looking much smaller with a scant cytoplasm and large nucleolus. However, even within this there was great levels of variation, often because of the rapid cell growth and division associated with this pluripotent cell type it may be hard to exactly distinguish a primary colony with complete accuracy. Therefore, what resembled primary iPSC colonies were excised and re-plated onto a fresh iMEF feeder layer for continued outgrowth and analysis.

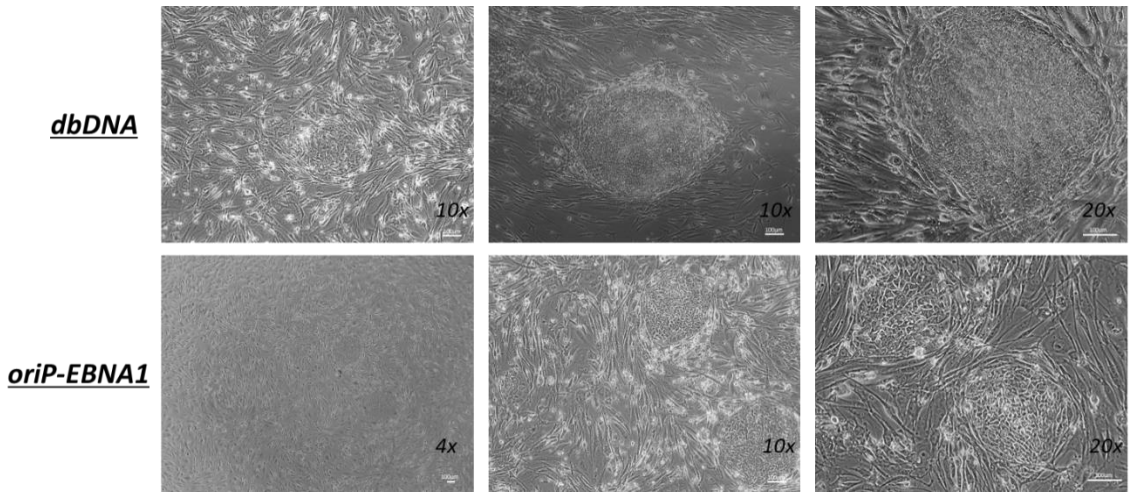



***Figure 25 – iPSCs developed from primary colonies which were isolated from reprogramming experiments incorporating both dbDNA and oriP-EBNA1 vectors in CLN3 mutated fibroblasts.***

Images depict subsequent iPSC colony growth following the isolation of primary colonies. The iPSCs for both vector types display a more typical iPSC phenotype and are shown up to passage 3.

The re-plated primary colonies for both vector types continued to grow on fresh iMEF feeder layers and demonstrated a more typical iPSC phenotype as previously described (*Figure 25*). The iPSCs produced by the dbDNA system were routinely passaged, displaying a stable phenotype persisting above and beyond p10. The iPSCs produced by the *oriP*-EBNA1 system were likewise concurrently passaged and maintained but however, did not stabilise and had undergone unwanted spontaneous differentiation and loss of pluripotency by passage 5. This was a surprising result, it was established by other lab members that these fibroblasts were able to be reprogrammed by the *oriP*-EBNA1 system but that the iPSCs produced were likewise never stable. For the dbDNA vector to not only successfully produce iPSCs, but to produce more stable iPSCs this was a remarkable result. Duly, we aimed to then reprogram and produce iPSC lines from as many different fibroblast samples as possible, with the intention of gaining a greater insight into the true functionality and reproducibility of the dbDNA system in the context of iPSC development.

*Table 3 - Table depicting fibroblasts lines reprogrammed using the dbDNA system.*

| Fibroblast line | Neonatal/ Juvenile/ adult | Reprogrammed with dbDNA | Passage (p) | Cryopreserved |
|---|---|---|---|---|
| Control #1 | Neonatal | Yes | p32 | Yes |
| Control #2 | Juvenile | Yes | p16 | Yes |
| Control #3 | Juvenile | Yes | p9 | Yes |
| Control #4 | Adult | No | n/a | n/a |
| Batten disease (CLN3-417PA) | Juvenile | Yes | p14 | Yes |
| Batten disease (CLN3-417PB) | Juvenile | Yes | p8 | Yes |
| Batten disease (CLN3-338PA) | Juvenile | Yes | p7 | No |
| Batten disease (CLN6-423PA) | Juvenile | Yes | p13 | Yes |
| Batten disease (CLN6-476PA) | Juvenile | Yes | p8 | Yes |
| Batten disease (CLN6-501PA) | Juvenile | Yes | p16 | Yes |
| Batten disease (CLN7-474PA) | Juvenile | Yes | p12 | Yes |
| Batten disease (CLN7-466) | Juvenile | Yes | P7 | Yes |

The dbDNA system successfully reprogrammed and produced iPSC lines from a total of 11 different fibroblast samples from both control and *Batten disease* origin (*Table 3*). The dbDNA system displayed a clear robustness within its functionality in respect to its ability to reliably reprogram and produce iPSCs from fibroblasts isolated from individuals of different ages and expressing different genetic mutations. Such iPSCs were moreover able to persist and maintain their stability with stocks being cryopreserved. Additionally, the dbDNA system also successfully produced iPSCs from fibroblasts that were previously refractory to reprogramming with the *oriP*-EBNA1 system. The *oriP*-EBNA1 system could produce primary/early passage colonies, however, the cells would spontaneously differentiate leaving nothing of a pluripotent

phenotype behind. While the dbDNA system was able to reliably produce much later passage iPSCs from the same fibroblast samples.

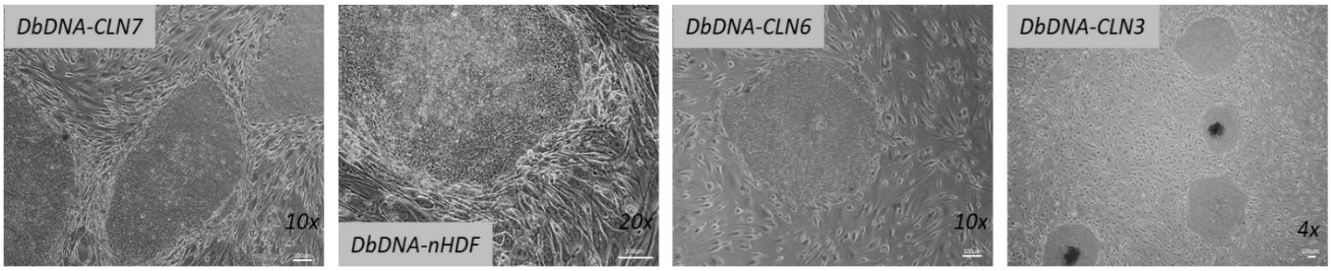



***Figure 26 - Example images from some of the iPSC lines produced using the dbDNA reprogramming vector.***

In its entirety, the dbDNA system displayed a great success in terms of its functionality in relation to the production of iPSCs and a robustness in terms of the fibroblasts it has successfully produced stabilised iPSC lines from (*Figure 26*). This was an interesting and successful result in terms of the dbDNA systems reliability and success where the *oriP*-EBNA1 vector had clearly fallen short. Such reproducible success led us to begin experiments with the aim of quantifying the dbDNA systems functional capacity in relation to the *oriP*-EBNA1 system in terms of its efficiency of iPSC colony production. An alkaline phosphatase stain (AP stain) was commonly utilised and critical to the quantification of the systems relative functionality. An AP stain can be used to determine the presence of *bona-fide* primary iPSC colonies, this number can then be calculated in proportion to the number of fibroblasts initially seeded onto the iMEF feeder layer to provide a percentage reprogramming efficiency.

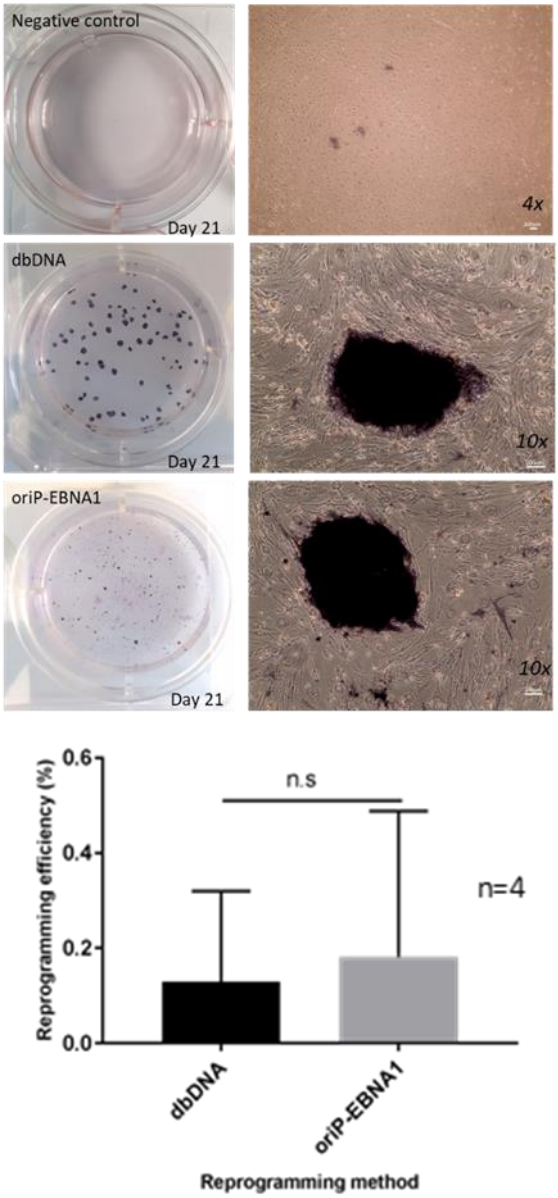



***Figure 27 - Alkaline Phosphatase (AP) stain for reprogramming experiments carried out using the dbDNA and oriP-EBNA1 systems.***

Image depicts typical result following an AP stain which specifically interacts with pluripotent colonies. The efficiency was calculated as the number of colonies formed per the total number of fibroblasts seeded. The bar chart represents the reprogramming efficiency as a percentage of colonies formed per fibroblasts seeded. The efficiency was derived from n=4 experiments. Error bars represent SEM of each vectors biological replicate. The data was subjected to a Kolmogorov-Smirnov test for normality with the p-value being >0.05 suggesting non-normal distribution. As such, a Mann-Whitney U-test was carried out per timepoint. A p-value of 0.78 was determined.

The dbDNA reprogramming system was demonstrated to function with a reprogramming efficiency that was not significantly different to that of the *oriP*-EBNA1 system (*Figure 27).* This was an interesting insight for two distinct reasons. Firstly, the

dbDNA system was a transient system with no scaffolding functionality as the *oriP*-EBNA1 system does. This, in theory, provides the dbDNA system with a much greater transiency in terms of its expression longevity than the *oriP*-EBNA1 system. In theory, this should result in the far reduced functionality of the dbDNA system in terms of iPSC development in relation to its scaffolding counterpart, but it doesn't. Secondly, the *shp53* transgene was incorporated within the *oriP*-EBNA1 system with the aim to improve the functional efficiency of the reprogramming process using the vector. This was in relation to *p53* and its function in the maintenance of cell cycle integrity. Cells exhibiting DNA damage would hit cell cycle checkpoints where p53, in synergy with other factors, could initiate apoptosis within cells that are too unstable to progress through such checkpoints. This means that potential cells that would have been reprogrammed instead, potentially, are not. Therefore, the short-hairpin RNA (shRNA) for the *guardian of the genome* was incorporated to improve the reprogramming efficiency. The dbDNA system as previously mentioned, does not incorporate this p53 shRNA. Yet, clearly, the system still not only has a very similar functional capacity to that of the *oriP*-EBNA1 system in terms or reprogramming but, can also reprogram iPSCs with a greater level of success than its counterpart too (*Figure 28*).

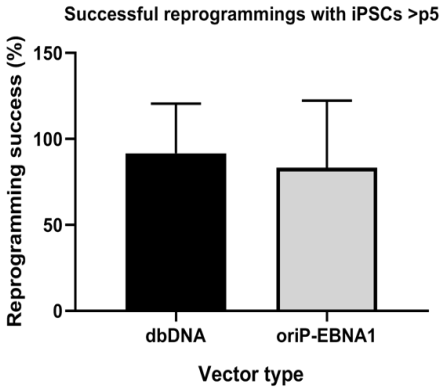



***Figure 28 – Success rate of reprogramming for both oriP-EBNA1 and dbDNA vectors in fibroblasts.***

The graph depicts the successful rate of reprogramming experiments carried out using both vector types. Successful primary colony production persisting beyond p5 was deemed successful. Bar chart representative of n=12 experiments. Error bars representative of SEM of reprogramming experiments for both vector types.

This greater level of success in reprogramming was due to the ability of the dbDNA system to reprogram and produce stabilised iPSC lines from fibroblasts that the *oriP*-

EBNA1 system did not have the same level of success with (*Figure 29*). The results evidenced that the dbDNA system, despite its transient design, had a functional equivalence as well as a robustness which surpassed that of the *oriP*-EBNA1 system.

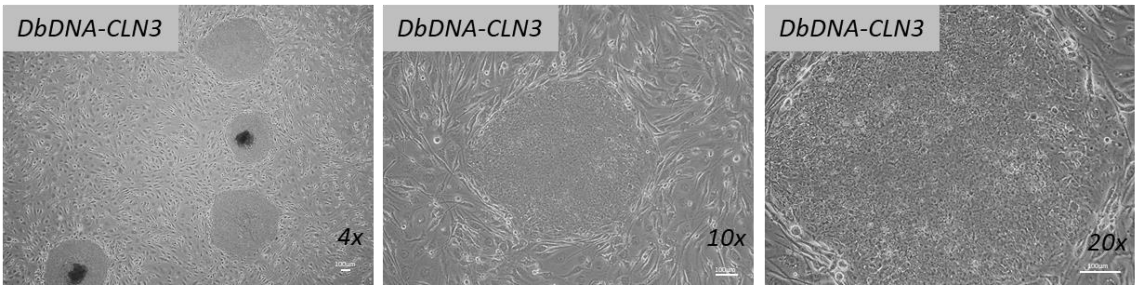



***Figure 29 - Representative images of the successful reprogramming of CLN3 (417PA) using the dbDNA system.***

Images depicting the stabilized iPSC line produced from the reprogramming of previously refractory primary fibroblasts using the dbDNA system. Images depict n=1 experiment.

In so far, the data has evidenced that the dbDNA vector has produced stable iPSC lines from a number of fibroblast sources. This was unique in the sense that the dbDNA system was designed as a transient system. Transient systems, such as plasmid, have previously displayed minimal functional success in terms of reprogramming and iPSC development. In turn, they have had to be modified to contain the *oriP*-EBNA1 sequence to improve the vectors longevity and ultimately its functionality within this process. It is, therefore, within reason to consider the results obtained using the dbDNA system as being extraordinary. In order to demonstrate the unique nature of the ability of the dbDNA system to produce iPSCs from fibroblasts, a large-scale reprogramming experiment was carried out. This employed the *oriP*-EBNA1 plasmid, the dbDNA system and a plasmid devoid of the *oriP*-EBNA1 system. The aim would be to demonstrate that the *oriP*-EBNA1 system was a necessary requirement to produce iPSCs when using plasmid. But, that the dbDNA system can be demonstrated to function independent of any scaffolding or tethering capacity.

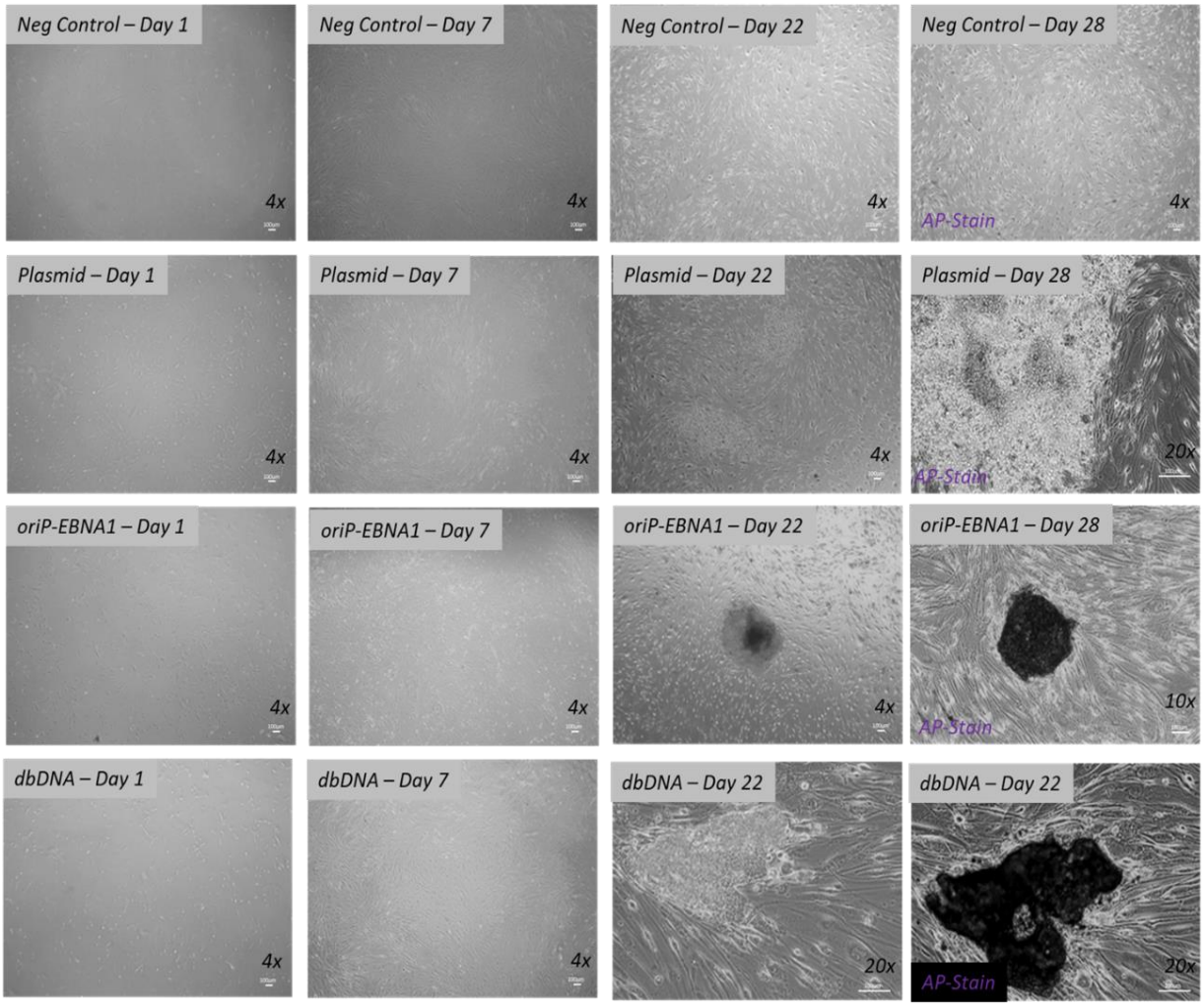



***Figure 30 – A reprogramming experiment to determine the novely of oriP-EBNA1-free iPSC production using the dbDNA system.***

A reprogramming experiment including a negative control (fibroblasts nucleofected with a GFP expressing plasmid), reprogramming plasmid (without oriP-EBNA1), oriP-EBNA1 reprogramming plasmid and dbDNA reprogramming vectors. All reprogramming experiments were carried out concurrently until on Day 28, the cells underwent AP staining to demonstrate the presence of any primary iPSC colonies. The negative control and the plasmid displayed no AP positive colonies. While the oriP-EBNA1 and dbDNA reprogrammed cells did display AP positive colonies. Images depict an n=1 for the experiment.

As *Figure 30* demonstrated, plasmid alone, that is without the *oriP*-EBNA1 system, was incapable of inducing a pluripotent phenotype within fibroblasts resulting in no positive alkaline phosphatase staining. This highlights even within our own reprogramming protocol, owing to a single transfection, that plasmid devoid of the *oriP*-EBNA1 system was incompatible with the reprogramming process and for plasmid to achieve any success in the induction of pluripotency in somatic cells that the accessory

EBNA1 system has to be included. This helped to substantiate the novel nature of the efficient and reproducible functionality of the dbDNA system within this same process.

### 2.1.4. **Pluripotency characterization of dbDNA/*oriP*-EBNA1 produced iPSCs:**

The dbDNA system has been able to produce what phenotypically resemble primary iPSC colonies, which when isolated and continually cultured resembled the iPSCs produced using the *oriP*-EBNA1 system. Having been able to produce such cells using the dbDNA system, we felt it important to ensure they had fully developed pluripotent properties. This means that the cells are expressing genes which are critical to the maintenance of the cells perpetual self-renewing phenotype, and that the cells can also differentiate to form phenotypes of all three germ lineages. The first experiment we aimed to carry out was to probe the cells for pluripotent markers using immunocytochemistry (ICC). Both transgene-specific (*OCT4*/*SOX2*) and endogenous (*NANOG*, *TRA-1-60*, *TRA-1-81*) protein markers were analysed. This was to ensure that not just pluripotent proteins which were present on the vector transgene were being upregulated, but that the cells have undergone complete reprogramming and were also constitutively expressing key pluripotent markers in an endogenous fashion. The ICC staining was also carried out using ESCs, a pluripotent gold standard, as a positive control.

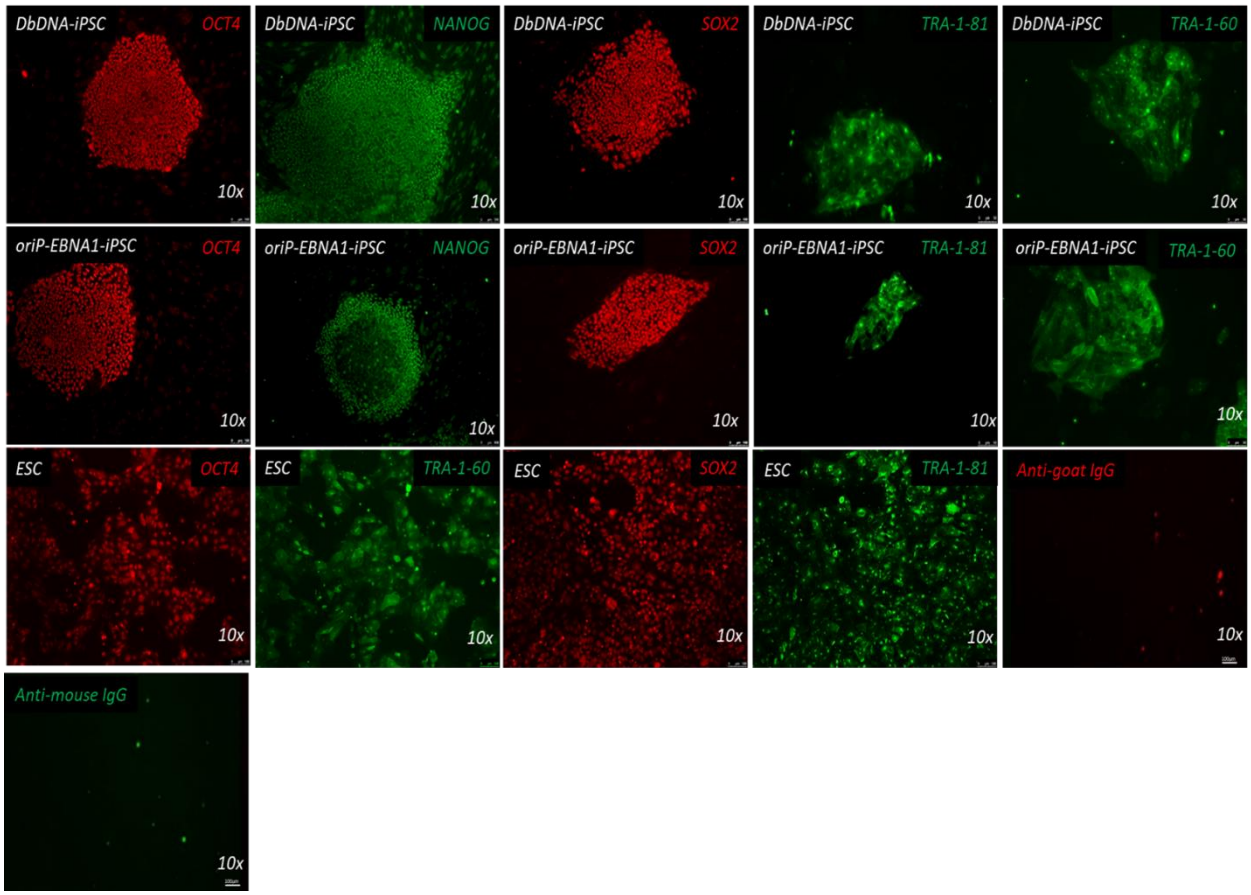



***Figure 31 - Immunocytochemical staining for key pluripotency markers in dbDNA-iPSCs and ESCs.***

Markers that are transgene specific (OCT4/SOX2) were probed alongside non-transgenic pluripotency factors (NANOG, TRA-1-81, TRA-1-60). The first row depicts dbDNA-iPSCs and the second oriP-EBNA1 iPSCs. Embryonic stem cells were stained as a positive control. Non-specific same species negative control igG antibodies were also employed. Images depict iPSCs produced all from the same fibroblast source in concurrent experiments between the two vectors. Scale bar represents 100um.

The ICC provided clear positive staining for key transgene-specific and endogenous pluripotency markers in both *oriP*-EBNA1 and ES cells alike (*Figure 31*). The cells produced by the dbDNA system also provided clear positive staining for all pluripotency markers. This clearly evidenced that the cells produced by the dbDNA system have developed a capacity to endogenously translate pluripotent mRNA into protein and were therefore pluripotent by nature. This was a great result and the first step to demonstrating that the cells produced by the dbDNA system were in fact *bona-fide* iPSCs. To further consolidate the results of the ICC, we wanted to carry out a Reverse Transcriptase-PCR (RT-PCR). The aim of this would be to demonstrate that the cells were actively transcribing pluripotent genes, thereby in combination with the ICC, the dataset would track such pluripotent genes through the central dogma (DNA

– RNA – Protein) of protein/gene expression. Moreover, the primers utilised within the experiments were designed only to amplify endogenously expressed genes by including part of the 3'UTR within the amplicon. This meant that we could now determine if the cells being analysed were endogenously transcribing pluripotent genes or if they were reliant on the expression of the transgene to maintain their pluripotent nature.

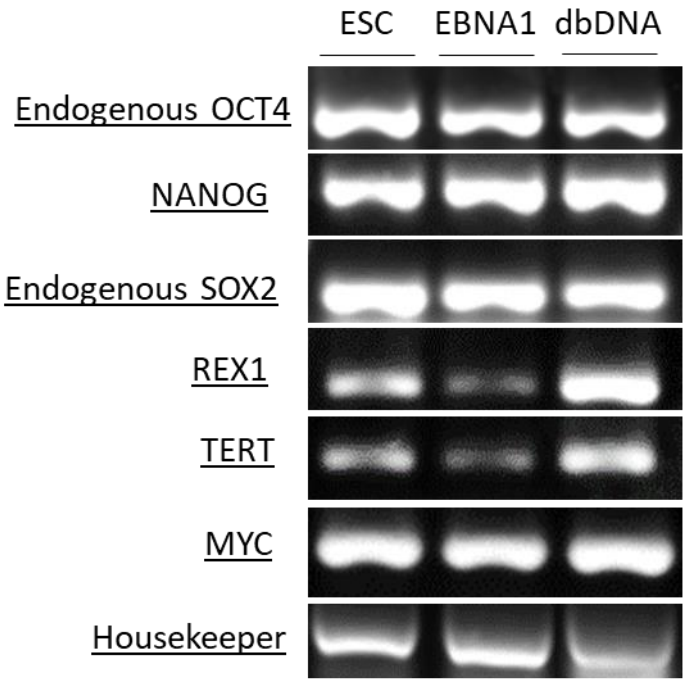



***Figure 32 - Reverse-transcription-PCR analyzing the transcription levels of key endogenous pluripotency markers.***

dbDNA and oriP-EBNA1-derived iPSCs alongside a positive control of ESCs were analyzed for pluripotent gene transcription. A housekeeper gene of RN18S1 was also employed for all 3 cell types too. The experiment represents a single iPSC line from the same fibroblast source for both dbDNA and oriP-EBNA1.

The RT-PCR consolidated the previous ICC data, demonstrating clear positive results for all probed endogenous pluripotency transcripts for both the dbDNA and the *oriP*-EBNA1 system alike (*Figure 32*). This evidenced that the cells produced using the dbDNA system resemble a pluripotent cell type both phenotypically and genotypically. The cells were actively transcribing key pluripotent markers endogenously and were also expressing pluripotent proteins. However, another key facet of iPSCs was that despite being able to maintain a pluripotent phenotype, the cells can also, following

external cues, differentiate to form cells of all 3 germ lineages (*Mesoderm, Endoderm* and *Ectoderm*). Thus far, we have been able to evidence that the cells produced by the dbDNA system demonstrated a perpetual pluripotent capacity. However, the next steps to confirming that the cells were truly iPSCs was to demonstrate that they can also form cells of the 3 germ lineages.

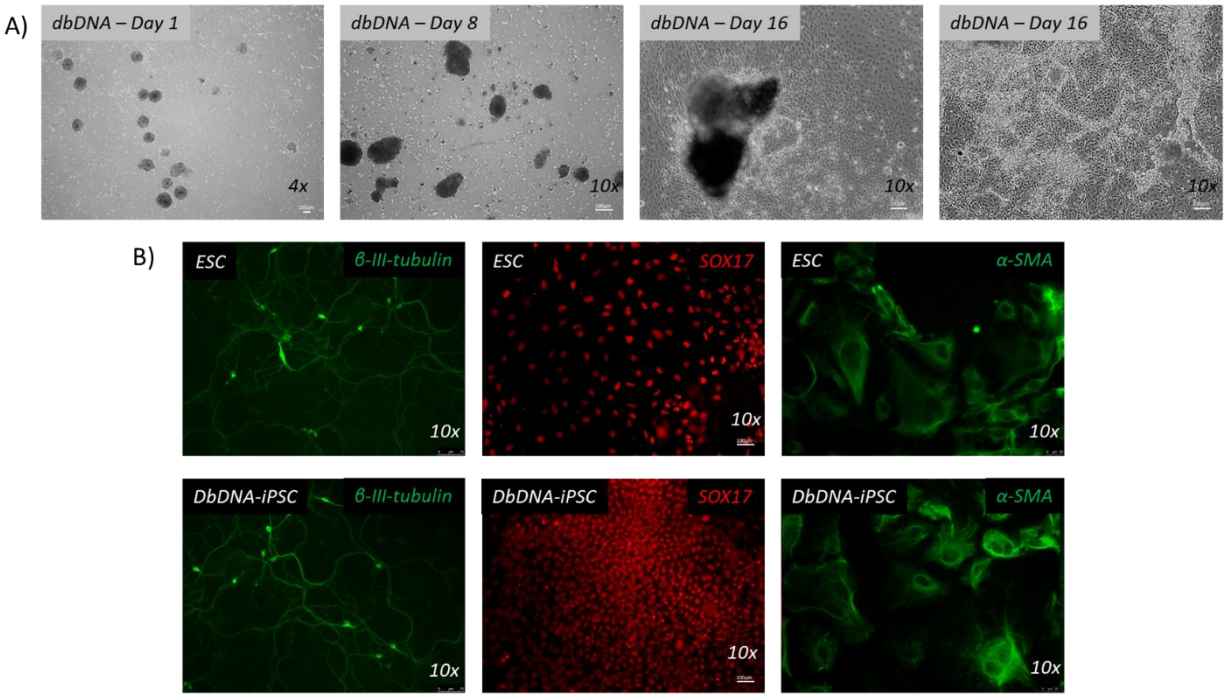



***Figure 33 – Embryoid body (EB) formation and subsequent ICC staining of the outgrowth for markers of all 3 germ lineages.***

iPSCs produced by the dbDNA system alongside the pluripotent gold standard ESC were analysed for their trilineage differentiation capacity. A) represents the phase contrast images depict different time-points from the experiment using dbDNA iPSCs. The first image, day 1, depicts EB formation using dbDNA iPSCs. While day 16 demonstrates the subsequent outgrowth from the same EBs which would ultimately undergo ICC analysis. The ICC stain depicted by B) demonstrates two rows of staining, ESC on the top row and dbDNA-iPSC on the bottom row. Then from left to right an ectodermal marker (B-III-tubulin), an endodermal marker (SOX17) and a mesodermal marker (alpha-smooth muscle actin).

The dbDNA system was able to generate embryoid bodies (EBs). EBs are a three-dimensional aggregation of pluripotent stem cells. Their formation is dependent upon homophilic binding between the cells mediated by E-Cadherin. The EBs were then re-plated and the outgrowth stained positively for markers of all 3 germ lineages using ICC (*Figure 33*).

The pluripotency ICC and RT-PCR experiments alongside the EB outgrowth and 3 germ layer staining demonstrated that the cells produced by the dbDNA system were *bona-fide* iPSCs. This was because they had been evidenced to exhibit the two most important features of iPSCs:

1. To perpetually self-renew and maintain an undifferentiated or *pluripotent* state.
2. To, on cue, differentiate to form cells of all 3 germ lineages.

Having clearly evidenced that our cells met such criteria, it was clear that the dbDNA system could generate iPSCs.

### 2.1.5. **Reprogramming experiment using Xenofree, cGMP compliant protocol:**

The results described so far provided proof-of-principle that dbDNA vectors can successfully be employed in iPSC production. The methodology employed so far however, utilised experimental constituents of animal origin. This can result in complications and inhibit the iPSCs potential downstream clinical use due to the possibility for the transmission of zoonotic disease to the cells. To reduce the likelihood of this and to aid the progression of this cellular technology into the clinic, *xenofree* methods of reprogramming have since been developed. These protocols employed experimental constituents which were completely animal free, thereby negating the use of specific medium constituents like *Foetal bovine serum* (FBS) and the iMEF feeder layer which was derived from mouse embryo. The rationale of the whole project was to incorporate the use of the novel dbDNA system in order to improve the translational potential of iPSCs for a therapeutic benefit. It was therefore critical to determine the vectors functional capacity within such clinically transferrable protocols. We therefore adopted a *xenofree* reprogramming method and employed both the dbDNA and oriP-EBNA1 vectors into this protocol.

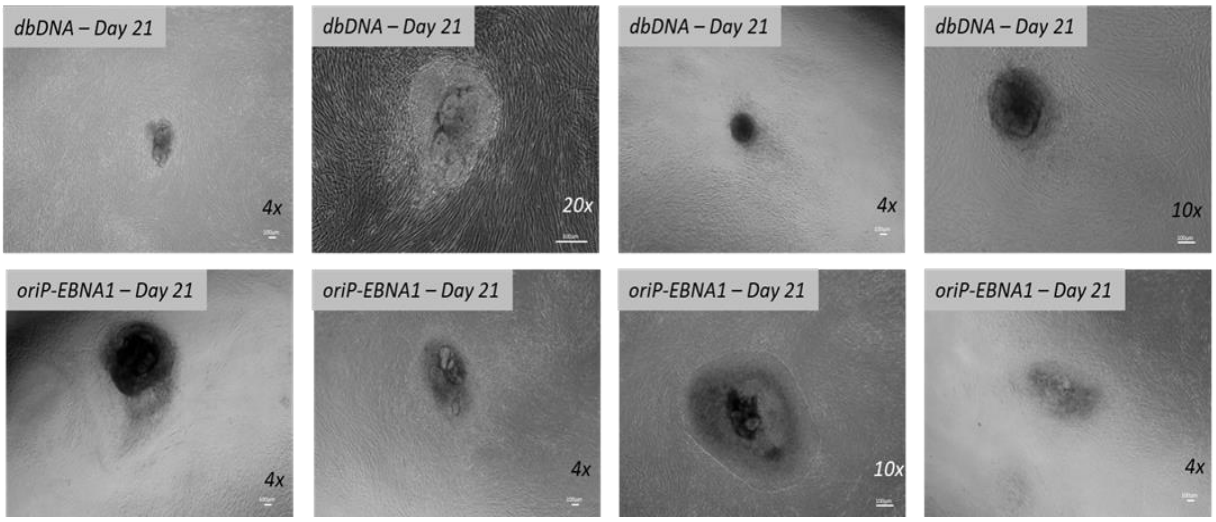



***Figure 34 - Reprogramming of hDFs using a xeno-free protocol employing both the dbDNA and oriP-EBNA1 vectors.***

Images depict primary colonies produced by both dbDNA (top row) and oriP-EBNA1 vectors when being incorporated into a xenofree protocol. The images depict a single experiment yielding multiple colonies for both vectors. The experiment was carried out concomitantly for both vectors and employing the same primary fibroblasts.

The dbDNA system was able to produce primary iPSC colonies when incorporated into the *xenofree* reprogramming protocol, as did the *oriP*-EBNA1 system (*Figure 34*). As mentioned, the system incorporated animal-free products with fibroblasts being cultured in *Essential 6*, before being moved onto the pluripotent *Essential 8* medium with vitronectin replacing the iMEF feeder layer. The *xenofree* method of reprogramming often results in a reduced reprogramming efficiency. Despite this, the dbDNA system (without additional transgenes such as the *shp53* sequence) was still able to function and produce iPSCs. This was an incredible result, incorporating such cells into a *xenofree* protocol provides proof-of-principle that a big hurdle, which many vectors fail to surpass without long-term efforts, can be overcome by the dbDNA system. From this, we decided to continue to culture the cells and carry out ICC staining for a key pluripotent marker, *SOX2*.

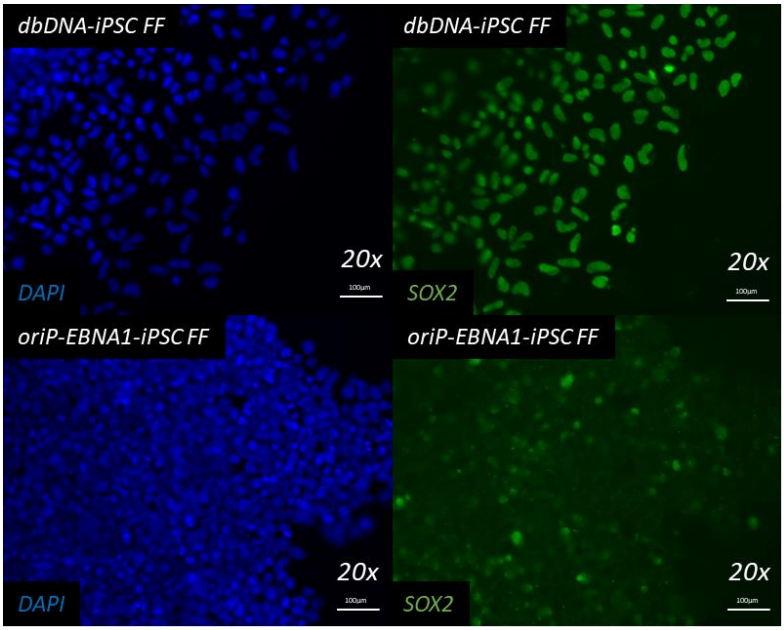



**Figure 35 - ICC staining of key pluripotency marker SOX2 on feeder free dbDNA/oriP-EBNA1 iPSCs produced using a xenofree compliant protocol/constituents.**

ICC staining for key pluripotent marker SOX2 on iPSCs produced using a xenofree protocol incorporating both the *oriP*-EBNA1 and dbDNA vectors. Scale bar represents 100µm.

The ICC provided positive staining of cells produced using both the dbDNA and *oriP*-EBNA1 system (*Figure 35*). This suggested that primary colonies produced by both vector types were pluripotent by nature. This was an incredibly exciting result and demonstrated that the dbDNA system was quite robust in terms of its reprogramming potential. This result then drove us to begin to question what other cell types can potentially be reprogrammed using this system outside of fibroblast biopsies.

### 2.1.6. **Reprogramming blood and urine-derived cells:**

Having demonstrated that the dbDNA system can reprogram dermal fibroblasts using both proof-of-principle and *xenofree* protocols, we wanted to next determine the novel vectors ability to reprogram different primary cell types. We applied the dbDNA system to attempt to re-constitute pluripotency in both urine-derived and peripheral blood. Firstly, urine samples were taken and the process of cellular isolation, outgrowth and maintenance was undertaken.

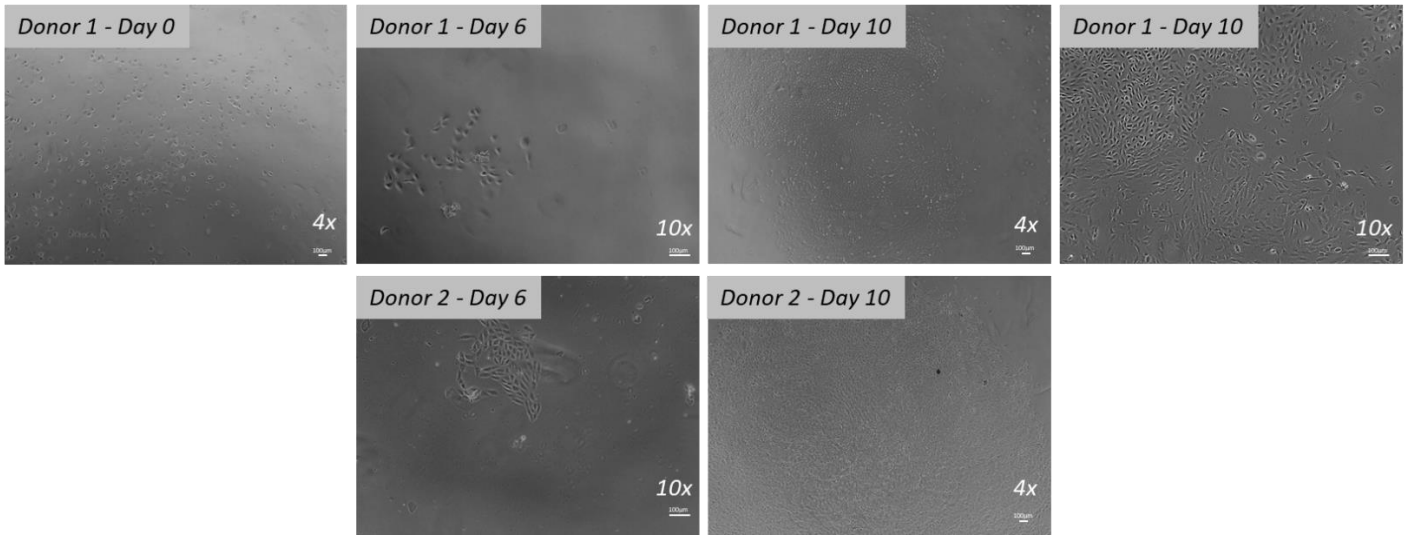



***Figure 36 - The isolation and amplification of urine-derived cells from two different donors.***

The phase images represent the process of isolating and culturing cells isolate from urine samples. The cells are isolated via centrifugation and continually cultured with daily medium changes before the cells begin to develop and grow out. By day 10 the cells should be confluent and ready to passage or reprogram. Scale bar represents 100µM.

The process of cellular isolation and expansion was fairly time consuming and laborious. The initial isolation protocol was lengthy with additional care being taken which minimised the increased risk of cellular infection. Once plated into medium, the cells required daily feeding, with populations beginning to appear around day 6 before becoming confluent 10 days post initial isolation and plating *(Figure 36*). Once confluent, stocks were developed and the remaining cells were incorporated into a reprogramming protocol utilising both the dbDNA and *oriP*-EBNA1 vectors.

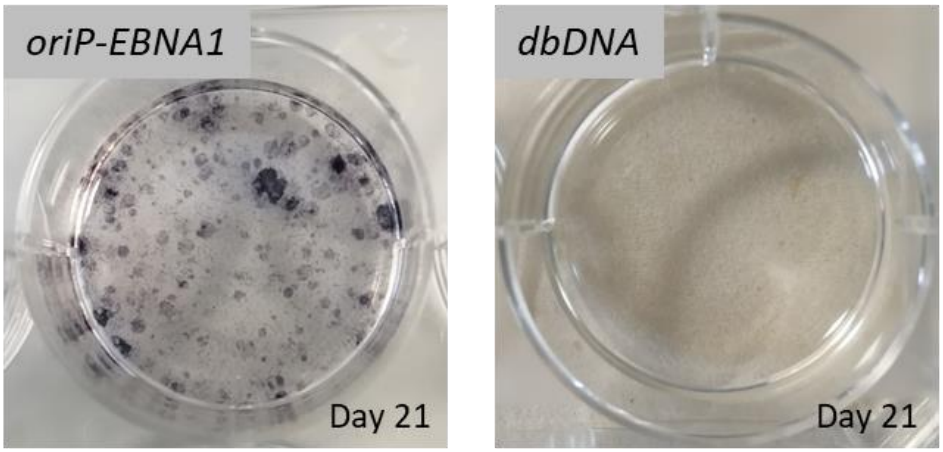



***Figure 37 - AP stain of primary iPSC colonies produced using both the oriP-EBNA1 and dbDNA reprogramming vectors in urine-derived cell types.***

Urine-derived cells were subjected to reprogramming using the oriP-EBNA1 and dbDNA systems. The images depict the cells 21 days post nucleofection having undergone an AP stain to determine the presence of pluripotent colonies. Images represent a single experiment carried out concomitantly.

The reprogramming of urine-derived cells yielded pluripotent colonies for the *oriP*-EBNA1 vector and not the dbDNA system (*Figure 37*). Having attempted the protocol several times, it was only successful on a single occasion. This was interesting and suggested maybe a technical issue rather than a potential mechanistic failure, however, more work is required to elucidate this. Subsequently, both vector types were also applied to the reprogramming of adherent peripheral blood mononuclear cells (PBMCs). This was a more invasive method of cellular isolation than isolating urine-derived cells but was considerably less invasive than a biopsy to isolate fibroblasts. The PBMCs were isolated from the buffy coat of a fractionated peripheral blood sample. The initial isolation of the cells and their subsequent nucleofection were all to be carried out immediately following the blood sample being taken. The first phase was therefore again incredibly time consuming and labour intensive.

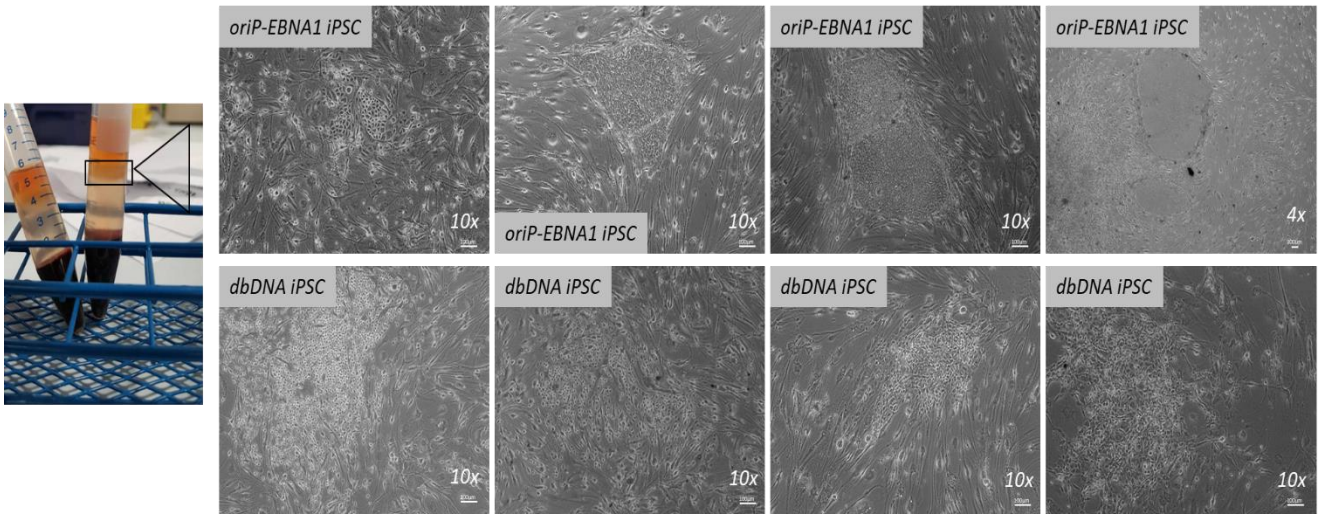



***Figure 38 - Reprogramming of PBMCs isolated from peripheral blood using both the dbDNA and oriP-EBNA1 system.***

Images depict the buffy coat from which the PBMCs are isolated prior to nucleofection and reprogramming. The phase images demonstrate the successful primary colony production using the oriP-EBNA1 system. The dbDNA system was unsuccessful in primary colony formation, however the cells produced from the experiment using the vector are documented. Scale bar represents 100µM.

The reprogramming of PBMCs was only successful using the *oriP*-EBNA1 vector, with the dbDNA system producing nothing of a primary colony phenotype (*Figure 38*). The dbDNA system resulted in the formation of an '*in-between*' phenotype not being a fully-fledged primary colony but different in phenotype to adherent PBMCs. Several titrations were then undertaken, different PBMC cell numbers were applied to the protocol alongside different concentrations of the dbDNA vector but with neither demonstrating any success. The primary colonies produced using the *oriP*-EBNA1 system were taken for pluripotency staining using ICC to determine the presence of key pluripotency proteins.

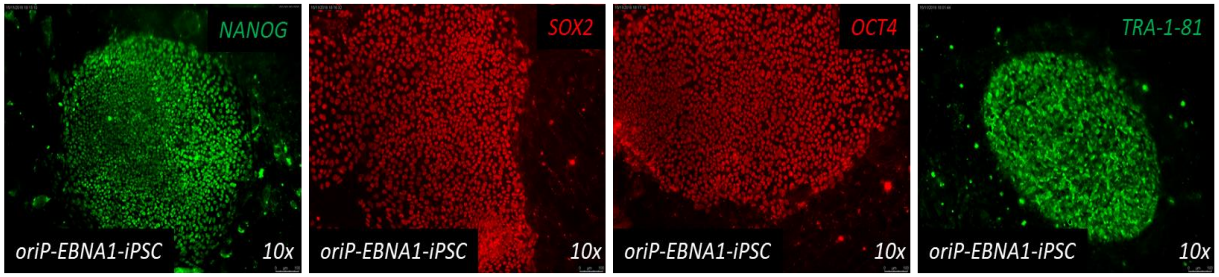



***Figure 39 - ICC staining for key pluripotency markers on iPSCs produced from reprogramming PBMCs using the oriP-EBNA1 vector system.***

Images depict the staining of iPSC colonies produced by the oriP-EBNA1 vector from adherent PBMCs. The cells were stained for key endogenous and transgene-present pluripotent markers. Scale bar represents 100µm.

The oriP-EBNA1 produced iPSCs were clearly pluripotent by nature, demonstrating positive ICC staining of key endogenous and transgene specific pluripotency markers (*Figure 39*). Titrations incorporating the dbDNA system within the PBMC reprogramming protocol were again undertaken, still with no further success. That said, the development of the in-between phenotype by the dbDNA system does suggest however that the reprogramming of this cell type isn't infeasible with this vector and could still be possible with future efforts.

### 2.1.7. Microarray analysis of dbDNA/*oriP*-EBNA1 produced iPSCs.

Following the production of the novel dbDNA iPSCs and the exceptional circumstances in which the vector had been able to function, led us to want to begin to investigate how the vector of choice was received by each respective host cell. The dbDNA system, as previously mentioned, was a clinical grade vector which omits the inclusion of any bacterial DNA, the EBNA1 sequence/protein and the *shp53* transgene. I subsequently wanted to investigate did an exclusion of such sequences from the dbDNA vector confer any benefit to the transfected host cell? A microarray analysis was determined to be suitable for the determination of such differences between both cell types on a global scale. An experiment was therefore designed which would incorporate iPSCs produced by both the dbDNA and *oriP*-EBNA1 systems. The iPSCs were produced concurrently, were at identical passages and produced from the same primary fibroblast cultures. Moreover, the primary fibroblasts were also included within the experimental design alongside the pluripotent '*gold standard*' ES cells. The experimental design helped us to substantiate the degree by which each iPS cell type had progressed away from its parental fibroblast and subsequently how closely it then aligned with the pluripotent ES cell type. Moreover, it allowed a direct analysis and comparison of the two iPSC lines, elucidating any systemic vector-mediated effects. Microarray analysis was carried kindly carried out at DKFZ German Cancer research Centre, Heidelberg, Germany. The Illumina HT12 beadchip system was used to analyse global gene expression in my samples. RNA was isolated from each of the biological triplicates used in the experiment before being DNase treated. This was a direct hybridization microarray, whereby 750ng of biotinylated cRNA isolated from the different experimental conditions was applied to the chip and hybridized for 17 hours. The HT12 system was able to determine the expression of over 47,000 different probes spanning the whole genome. Once completed, the data was processed and normalised by quantile normalisation using R studio. Subsequently, all probes and their expression values for each cell type was then provided to me.

From here, whole probed gene-sets were firstly analysed in an unbiased approach using *R-studio*.

*R-Studio* was initially used to produce a correlogram. A correlogram was a figure that would provide the correlation coefficient for each analysed cell type for designated probe expression values. Two identical cell types will have a correlation coefficient of '*1*' with divergences in probe expression between two cell types promoting a decline in this number, with two datasets of complete opposite probe expression profiles having a correlation coefficient of *'-1*'. I decided to produce a correlogram between all cell types analysed in the microarray across the entire probed transcriptome for each cell type.

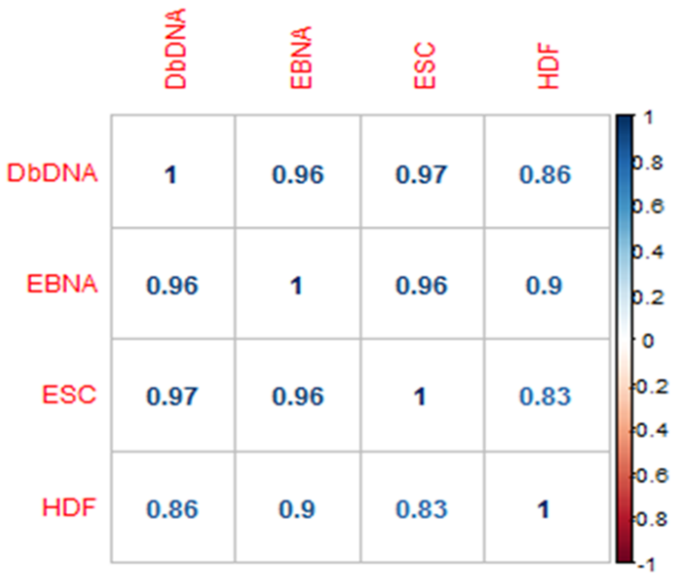



***Figure 40 - Correlogram depicting the correlation coefficient for each cell type across all gene probes included within the microarray analysis.***

A correlogram produced using R-Studio which provides the correlation coefficient of two analyzed cell types to determine similarities within their expression profiles. Two identical cell types with have a correlation coefficient of '*1*' with divergence between the expression profiles of two cell types driving this numerical value down. The correlogram details the coefficient values for each cell type analyzed within the microarray in comparison to itself and each other cell type also included in the experiment.

The analysis suggested a high degree of similarity in the expression profiles of all the pluripotent cell types incorporated in the experiment. The highest degree of similarity shared between 2 different cell types was between dbDNA iPSCs and ESCs. This was closely followed by EBNA iPSCs-ESCs and EBNA iPSCs-dbDNA iPSCs both of which demonstrated a coefficient value 0.01 lower than that in relation to that of Doggybone and ESCs (*Figure 40*). That said, the pluripotent cell types all share great similarity in

their coefficient values between one and other and were all most significantly divergent from that of the parental fibroblasts. ESCs had the most reduced coefficient value in comparison to the parental fibroblasts at '*0.83'* suggesting key differences within probed gene expression between these two cell types. dbDNA produced iPSCs had the second lowest coefficient value in respect to HDFs at '*0.86'*, yet *oriP*-EBNA1-produced iPSCs had a coefficient value of '0.90' in comparison to its parental fibroblasts; the highest of all the pluripotent cell types. This suggested a closer homology/reduced divergence between these two cell types in comparison to both dbDNA iPSCs and ESCs.

In order to attempt to elucidate the similarity of probed gene expression further, a *principle component analysis* (PCA) was undertaken. The aim of this analysis would be to take our complex gene expression dataset and represent it on a 2D graph with each cell type being plotted based on its relative variance to each other cell type. Therefore, two cell types with very distinct, differential probed gene expression within a *principle component* will be spaced further apart than two cell types that displayed a reduced level of variation within the same *principle component*. A precursor to a PCA was the production of a *scree plot*. A *scree plot* determined the percentage of complete variation captured within each *principle component*. For our experiment, we wanted to capture as close to 100% of the variation as possible to accurately represent the differences between each cell type in terms of their probed gene expression. Therefore, a *scree plot* determined the number of principle components that were required to be analysed within a PCA to capture and display the highest proportion of variation within our dataset.

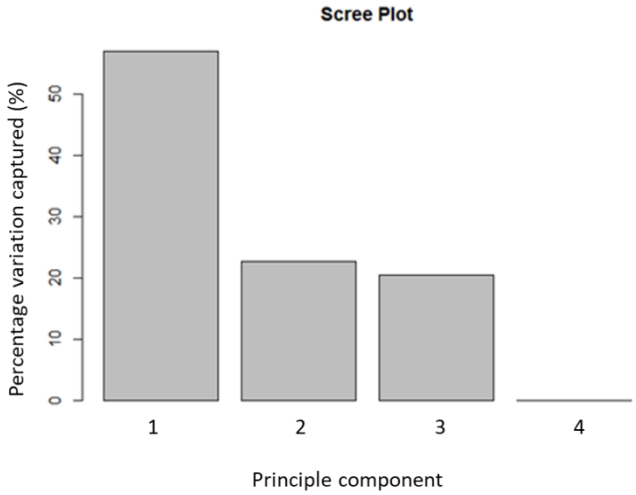



***Figure 41 - Scree plot identifying the level of variation captured within each principle component for the microarray dataset.***

A scree plot analysis was carried out to determine the level of variation captured within each principle component thereby dictating the number of components to analyze. The X-axis displays each principle component. The Y-axis, the percentage of variation captured within each principle component.

The scree plot demonstrated that in order to account for all the variation between each cell type in the microarray that 3 principle components must be analysed (*Figure 41*). As such, a PCA was carried out across PC1, PC2 & PC3.

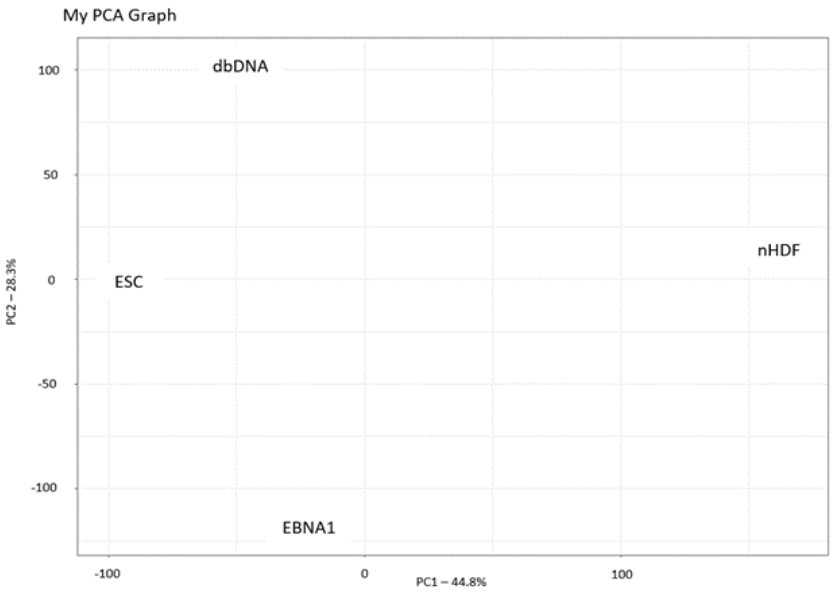



***Figure 42 - PCA plotting variation captured between PCA1 and PCA2 across the entire microarray dataset.***

The PCA will present all the variation between the datasets captured in PC1 and PC2 and present them on a 2D graph. The label of each cell type represents the averaged position for that cell type based on variation within its own probed gene expression in comparison to the expression values of each of the other analyzed cell types too. Therefore, the distance between 2 plots on a single axis was proportionate to the variability between the expression profiles of those plots. The further apart the plots are, the greater the divergence of probe expression. X-axis represents PC1 and Y-axis represents PC2.

The PCA dataset, representing PCA 1 and PCA 2 encapsulated the majority of variation within the microarray. This analysis presented that both dbDNA iPSCs and EBNA1 iPSCs were much more divergent in terms of their probe expression than the fibroblast dataset, with both being more similar and therefore plotted more closely to ESCs (*Figure 42*). The analysis also indicated that key probe expression differences existed between dbDNA and *oriP*-EBNA1 produced iPSCs; particularly in PC2. A second PCA was undertaken between PC1 & PC3 to account for the remaining variation.

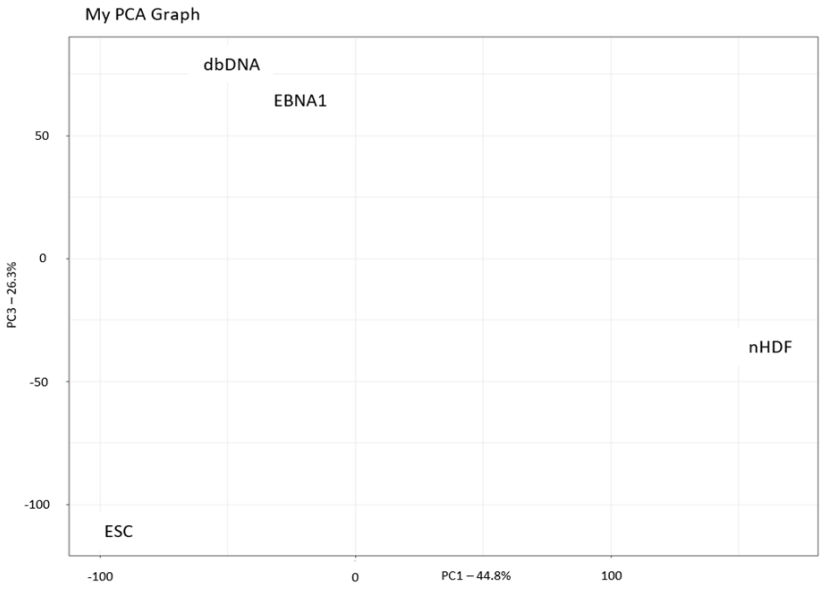



***Figure 43 - PCA plotting variation captured between PCA1 and PCA3 across the entire microarray dataset.***

The PCA will present all the variation between the datasets captured in PC1 and PC3 and present them on a 2D graph. The label of each cell type represents the averaged position for that cell type based on variation within its own probed gene expression in comparison to the expression values of each of the other analyzed cell types too. Therefore, the distance between 2 plots on a single axis was proportionate to the variability between the expression profiles of those plots. The further apart the plots are, the greater the divergence of probe expression. X-axis represents PC1 and Y-axis represents PC3.

The variation captured within PC3 conveyed a similar result to that of PC1 and PC2; that the probed gene expression in pluripotent cell types was more similar to one-and-other than that of the fibroblast samples (*Figure 43*). This was true for the two iPS cell types, yet PC3 did seemingly suggest a greater level of variation between the iPSC cell types and ESCs than previously captured in the two prior PCs. PC3 seemingly details a close similarity between *oriP*-EBNA1 and dbDNA produced iPSCs in terms of probe expression.

Yet, taken in its entirety the PCA demonstrated that differences exist in the expression profiles between all cell types analysed in the microarray. A key interest we wanted to explore was differences in probed gene expression specifically between dbDNA and oriP-EBNA1 produced iPSCs and what, if any, influence the vector used to produce such iPSCs was having in terms of manipulating differential probe expression.

From here, all the biological repeats from the microarray were averaged for each cell type to provide an average probe expression values for each cell type and probe. The dbDNA and *oriP*-EBNA1 probe expression values were then subjected to a *student t-test* analysis with a cut off *p-value* of ≤0.05 to indicate significantly expressed genes between the two cell types. This limits the possibility of spurious results to 5%. However, all probes with a significant *p-value* and they're relative expression values were taken forward and subjected to a *Benjamini-Hochberg* analysis to determine the *false discovery rate* (FDR). The FDR was presented as a *q-value*. This was an analogue of a *p-value* which had been subjected to multiple hypothesis testing. The inclusion of the FDR analysis reduced the possibility of a *type 1 error* (whereby there was an accidental rejection of a true null hypothesis for particular probes) and therefore limited the inclusion of *false positive* results within the dataset. Again, an FDR cut off *q-value* of ≤0.05 was utilised. All probes with a significant *q-value* (≤0.05) were taken forward. A fold change expression was calculated between the dbDNA and *oriP*-EBNA1 differentially expressed probes. Within many analyses, probe values with a fold change of ≥1.2 were deemed to be differentially expressed. However, to improve the stringency and possibility of detecting real change between the two cell types, only probes with a fold change ≥1.5 between dbDNA and *oriP*-EBNA1 iPSCs was taken for further analysis and determined to be differentially expressed.

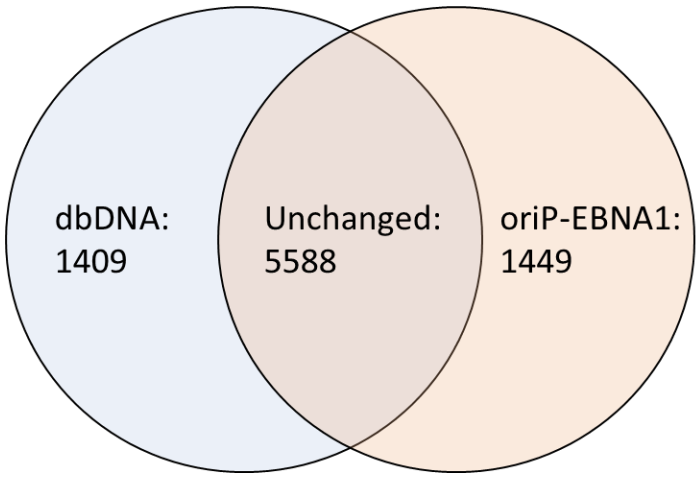



***Figure 44 - Venn diagram depicting the number of differentially expressed probes between dbDNA and oriP-EBNA1 datasets.***

The venn diagram depicts the number of differentially expressed probes between both dbDNA and oriP-EBNA1 produced iPSCs. The probe expression values for both cell types were subjected to multiple hypothesis testing prior to the fold change analysis – whereby only cells with a fold change ≥1.5 were taken as differentially expressed.

The analysis detailed that following the multiple hypothesis testing that 8443 probes were differentially expressed between *oriP*-EBNA1 iPSCs and dbDNA iPSCs to a significant degree. Following fold change calculations, of those 8443 probes, 1409 were over-represented in dbDNA produced iPSCs in comparison to *oriP*-EBNA1 iPSCs and 1449 were over-represented in oriP-EBNA1 in comparison to the dbDNA system (*Figure 44*). This may seem like a large proportion of probes for each vector. However, when considering the margin for variability within the reprogramming process, for example, in the ability of each vector to properly induce pluripotency within fibroblasts to the same degree alongside the effects any residual vector may be potentiating within the same cells, it was clear to see why such variation existed. To validate this finding further, a volcano plot was generated using *R-studio*. A volcano plot was a type of scatter plot which could quickly identify changes in large datasets. The volcano plot will combine a statistical significance with a magnitude of change. Therefore, incorporating our previous statistical principles, script was developed to produce a volcano plot.

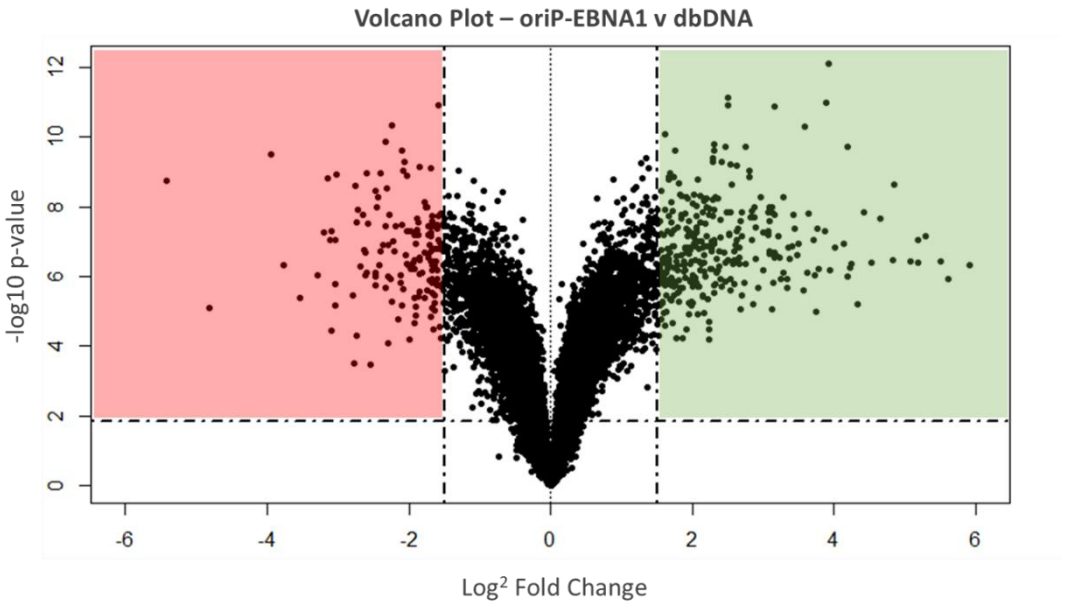



***Figure 45 - Volcano plot of the probes over/under-represented in oriP-EBNA1 iPSCs in comparison to dbDNA-iPSCs.***

Volcano plot represents a type of dot plot whereby each black dot represents an individual probe. With the combination of the statistical significant difference (≤0.05) and the magnitude of change (Fold change ≥1.5) we were able to isolate specific probes (dots) that were over-represented in oriP-EBNA1 iPSCs (green column) and probes that were over-represented in dbDNA produced iPSCs (red column).

The volcano plot result inferred that there was a number of significantly over-represented probes that were unique to both *oriP*-EBNA1 and dbDNA iPSCs in comparison to one-and-other (*Figure 45*). The *oriP*-EBNA1 iPSCs seemingly display a greater number of over-represented probes, helping to consolidate the previous numbers. The data thus far has clearly evidenced that there was key differences within the probed transcriptomics of iPSCs produced using the two different vector types. However, what biological processes the differentially expressed probes interact within was still unknown.

The first analysis that we wanted to undertake was to identify and analyse key pluripotency markers between all the cell types analysed in the microarray. This was carried out using a heatmap produced using *R-studio* incorporating all cell types analysed in the experiment.

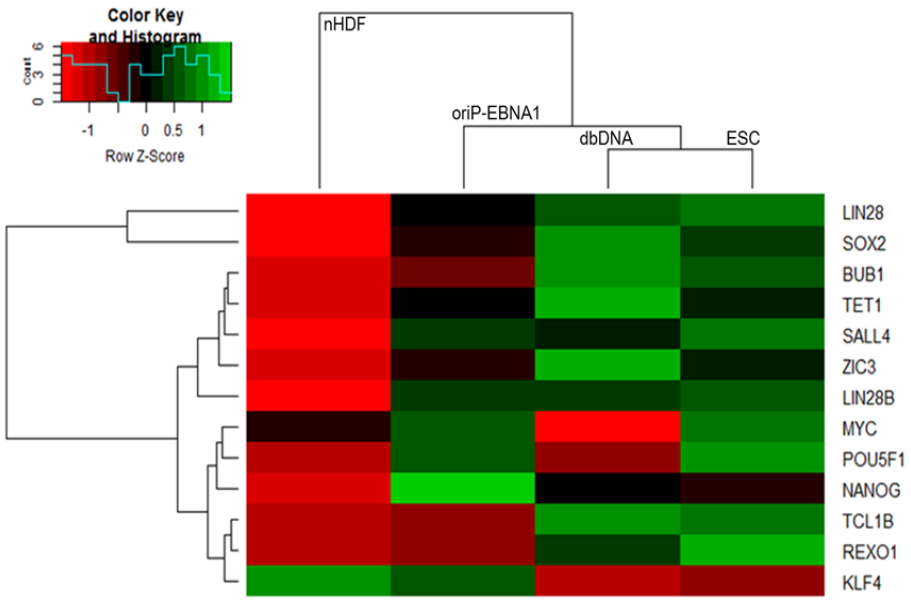



***Figure 46 - Heatmap demonstrating the relative expression of key pluripotency markers between nhDF, oriP-EBNA1-iPSCs, dbDNA-iPSCs and ESCs.***

Heatmap produced using R-studio depicting several pluripotency-specific probes across all cell types incorporated within the microarray. The hierarchical clustering amongst the cells determined which cell types probed gene expression pattern was most similar to one and other across such pluripotency probe sets and which were more divergent.

The hierarchical clustering tool within the heatmap allows visualization of pluripotency probe expression as a whole for each cell type and for the specific pluripotency probes incorporated in the analysis. The hierarchical clustering demonstrated on the whole that dbDNA produced iPSCs share more in common in terms of pluripotency probe expression with ESCs than *oriP*-EBNA1 produced iPSCs do with either cell type (*Figure 46*). However, the hierarchical clustering did suggest that while the probe expression values within dbDNA-produced iPSCs were more similar to that of ESCs, that *oriP*-EBNA1 produced iPSCs were still likewise very similar in their level of probe expression.

Following this, our next aim was to determine if the probes which were over-represented in *oriP*-EBNA1 produced iPSCs were occurring independently within this cell type and were therefore being influenced by the vector. I carried out a heatmap analysis incorporating the over-represented probes from *oriP*-EBNA1 iPSCs in comparison to dbDNA iPSCs. To provide the analysis with greater context, the expression values for the same probes from ESCs was also included. This assisted in

determining whether the over-represented probes within *oriP*-EBNA1 produced iPSCs were specific to that cell type only, or, if the expression pattern of such EBNA1 iPSCs was more similar to that of ESCs therefore indicating a failure of the behalf of the dbDNA system to faithfully induce gene expression within its own iPSCs.

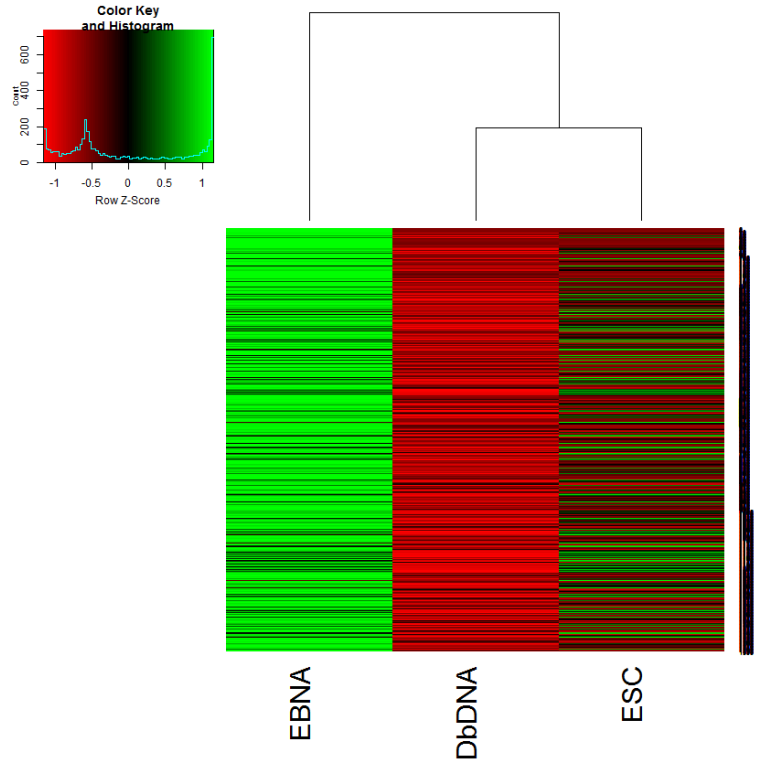



***Figure 47 - Heatmap comparison of probes that are over-represented in oriP-EBNA1 iPSCs in comparison with dbDNA iPSCs and ESCs.***

A heatmap produced using R-studio. Image depicts all the over-represented gene probes within oriP-EBNA1 iPSCs with green being highly expressed, black in-between and red a low level of expression. Hierarchical clustering also utilized to determine the similarities between cell types in respect to the gene sets.

The heatmap in conjunction with the hierarchical clustering evidences that, on the whole, dbDNA iPSCs share a more similar expression profile with that of ESCs in relation to *oriP*-EBNA1 over-represented probes than EBNA1-iPSCs (*Figure 47*). It was clear that, even though some individual genes do share a similar expression profile between EBNA1 iPSCs and ESCs, that ESCs generally didn't share similar express levels of most of the probes which were over-represented in *oriP*-EBNA1

iPSCs. Their expression levels instead were much lower and consistent with that of the dbDNA system. This therefore indicated that most of the probes that were over-represented in *oriP*-EBNA1 iPSCs, were done so in that cell type only. This suggested that the *oriP*-EBNA1 vector may be the root aetiological cause for the manipulation of some of the over-represented probes previously outlined. Following this exciting dataset, we wanted to determine what biological processes the *oriP*-EBNA1 specific probes were implicated within.

### 2.1.8. **OriP-EBNA1 over-represented probe analysis:**

Suggestions from the above result indicated that probes that were over-represented within *oriP*-EBNA1 iPSCs following a comparison with dbDNA iPSCs, were likewise over-represented in comparison to ESCs. The implications from this suggested that such probes were over-represented in *oriP*-EBNA1 produced iPSCs specifically and were therefore being manipulated in a vector-specific, pluripotency independent manner. To understand what effects the vector may be having on the cells, we aimed to determine what specific biological processes such over-represented probes were interacting with. Duly, several different software packages were utilised to determine transcription factor enrichment terms. The *Reactome* database was utilised to project significant genes onto the human genome, helping to elucidate interacting pathways within over-arching biological processes such as cell cycle, metabolism and immune function for example.

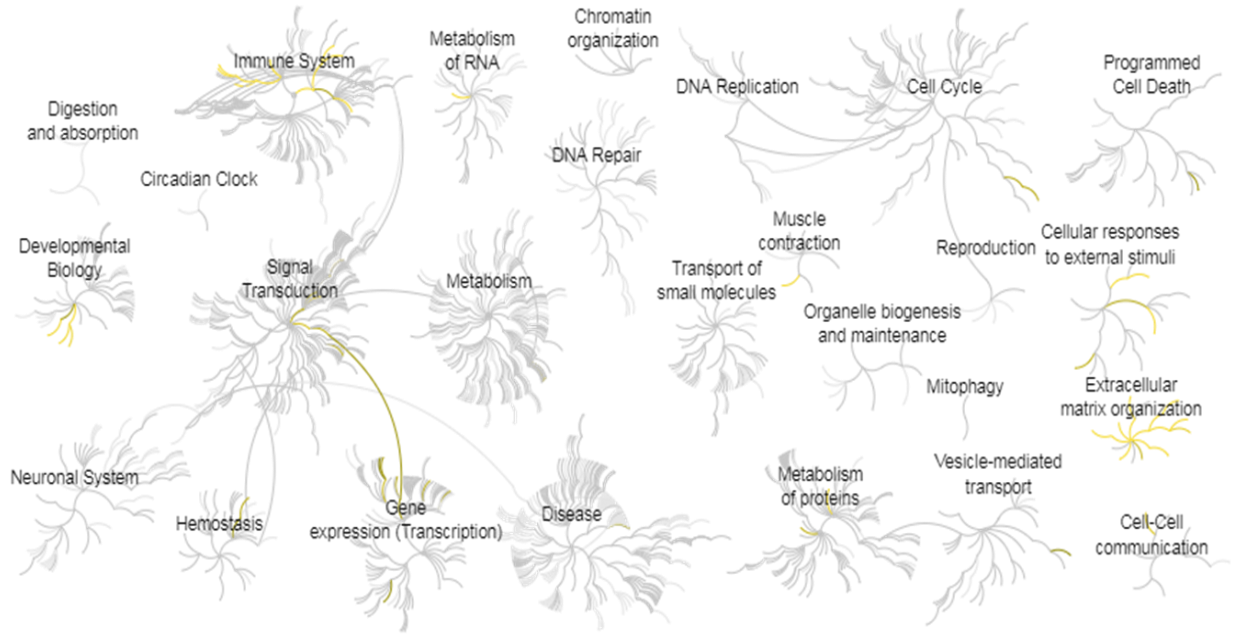



***Figure 48 - Projection of oriP-EBNA1 over-represented probes onto the human genome as determined by the Reactome database.***

oriP-EBNA1 over-represented probes were incorporated into the Reactome database which projects them onto different biological processes with which the probes most specifically align with. Each pathway has a focal point in the center with relating pathways being represented as branches protruding from this center point. Any pathways determined to be significantly interacting with the upregulated probes input into the system will be highlighted as yellow. FDR cut off <0.05 was employed.

The projections made by the Reactome database from the *oriP*-EBNA1 over-expressed probes had significant interactions within several different biological processes. Processes in relation to the *immune system*, *cellular responses to extracellular stimuli*, *extracellular matrix organisation* and *developmental biology* were but a few which were over-represented (*Figure 48*). The Reactome system was also able to quantify the number of probes which interacted with different biological processes to provide a clear understanding of which processes our probes were most commonly aligning in.

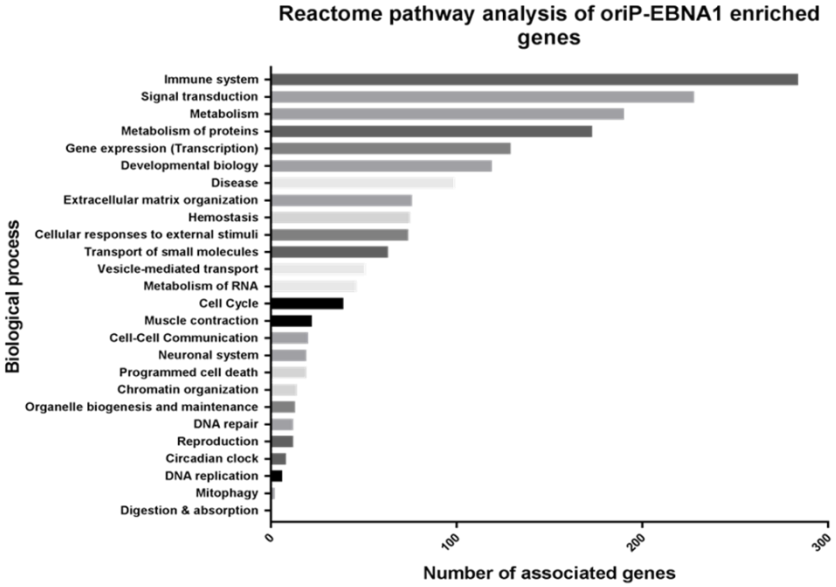



***Figure 49 - Quantification of the number of interacting probes within the main human pathways as projected by Reactome.***

The Reactome database will provide the number of probes, an inserted by ourselves into the database, interact with overarching human biological processes.

The quantification of the probes within different biological processes according to Reactome demonstrated that *immune system* processes was the most common pathway that the *oriP*-EBNA1 over-represented probes interacted within (*Figure 49*). A greater level of depth could be extracted from this dataset by investigating how the probes were interacting within such overarching pathways. In order to remove any bias from this analysis, biological processes with which the *oriP*-EBNA1 probes interacted were determined from 3 separate sources: *Reactome*, *Gene set enrichment analysis* (*GSEA*) and *Gene ontology* (*GO*). To begin with, the over-represented specific biological pathway for *oriP*-EBNA1 over-represented probes were analysed using *Reactome*.

| Pathway name | Entities found | Entities total | Entities pValue | Entities FDR | Reactions found | Reactions total | Reactions ratio |
|---|---|---|---|---|---|---|---|
| Endosomal/Vacuolar pathway | 60 | 82 | 1.11E-16 | 2.55E-14 | 4 | 4 | 3.25E-04 |
| ER-Phagosome pathway | 66 | 165 | 1.11E-16 | 2.55E-14 | 10 | 10 | 8.12E-04 |
| Antigen Presentation: Folding, assembly and peptide loading of class I MHC | 61 | 102 | 1.11E-16 | 2.55E-14 | 15 | 16 | 0.0013 |
| Antigen processing-Cross presentation | 69 | 187 | 1.11E-16 | 2.55E-14 | 17 | 23 | 0.001869 |
| Interferon gamma signalling | 88 | 250 | 1.11E-16 | 2.55E-14 | 11 | 15 | 0.001219 |
| Interferon Signalling | 106 | 392 | 1.11E-16 | 2.55E-14 | 41 | 66 | 0.005362 |
| Interferon alpha/beta signalling | 85 | 184 | 1.11E-16 | 2.55E-14 | 11 | 20 | 0.001625 |
| Extracellular matrix organization | 77 | 329 | 5.87E-10 | 1.18E-07 | 230 | 318 | 0.025837 |
| Cytokine Signalling in Immune system | 206 | 1261 | 2.17E-09 | 3.88E-07 | 277 | 699 | 0.056792 |
| Assembly of collagen fibrils and other multimeric structures | 25 | 67 | 1.70E-07 | 2.74E-05 | 25 | 26 | 0.002112 |
| Integrin cell surface interactions | 28 | 86 | 4.78E-07 | 6.98E-05 | 40 | 54 | 0.004387 |
| Transcriptional regulation of pluripotent stem cells | 19 | 45 | 8.45E-07 | 1.13E-04 | 34 | 35 | 0.002844 |
| Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell | 65 | 316 | 1.12E-06 | 1.33E-04 | 18 | 43 | 0.003494 |
| Elastic fibre formation | 19 | 46 | 1.16E-06 | 1.33E-04 | 14 | 17 | 0.001381 |
| Collagen formation | 29 | 104 | 5.94E-06 | 6.36E-04 | 58 | 77 | 0.006256 |
| Degradation of the extracellular matrix | 36 | 148 | 9.57E-06 | 9.57E-04 | 65 | 105 | 0.008531 |
| Class I MHC mediated antigen processing & presentation | 82 | 465 | 1.55E-05 | 0.001472 | 38 | 48 | 0.0039 |
| Molecules associated with elastic fibres | 15 | 38 | 2.50E-05 | 0.002222 | 8 | 10 | 8.12E-04 |
| ECM proteoglycans | 23 | 79 | 2.74E-05 | 0.002327 | 20 | 23 | 0.001869 |
| Non-integrin membrane-ECM interactions | 19 | 61 | 5.55E-05 | 0.004441 | 14 | 22 | 0.001787 |

***Figure 50 - Over-represented pathways determined by the Reactome database - ordered from the most significant p-value.***

Specific interacting pathways that the oriP-EBNA1 over-represented gene set interacts with as projected by the Reactome database. Reactome provides information on the number of entities or probes which interact within this process and likewise the entities p-value/FDR which relates to the specificity of the projections too. The processes are ordered from most significant p-value down.

The analysis using *Reactome* demonstrated a number of pathways associated with the immune system which were aligned with a high level of specificity (*Figure 50*). *Interferon signalling* including both *interferon alpha/beta signalling* and *interferon gamma signalling* were highlighted, alongside pathways associated with *MHC expression* and *Cytokine signalling in the immune system*. Overall, the *Reactome* database implicated the EBNA1-iPSC over-represented probes with having a largely pro-inflammatory signature.

Secondly, enrichment analysis was also undertaken using the *MSigDB* function from *GSEA*. This function provided information on *hallmark genes* which summarized and represented specific, well-defined biological processes generated via overlaps between gene sets within the *MSigDB* system. Subsequently, the analysis can provide a *p-value* for each hallmark process which demonstrates a measure of how significant the changes were for each given probe set – the higher the absolute value of the

statistic, the greater its significance. GSEA also provides k/K values, whereby k = the number of genes in the query set and K = the number of genes in the *MSigDB* database. This can therefore provide information on the number of probes interacting within a specific pathway. Finally, a *q-value* was provided which was an FDR analogue of the p-value after correction for multiple hypothesis testing and again reduces the possibility of including false positive results.

| Gene Set Name | Genes in Gene Set (K) | #Genes in Overlap (k) | k/K | p-value | FDR q-value |
|---|---|---|---|---|---|
| EPITHELIAL_MESENCHYMAL_TRANSITION | 200 | 95 | 0.475 | 4.44E-86 | 2.22E-84 |
| HYPOXIA | 200 | 58 | 0.29 | 2.31E-38 | 5.77E-37 |
| TNFA_SIGNALING_VIA_NFKB | 200 | 57 | 0.285 | 2.91E-37 | 4.84E-36 |
| INTERFERON_GAMMA_RESPONSE | 200 | 51 | 0.255 | 6.79E-31 | 8.49E-30 |
| APOPTOSIS | 161 | 40 | 0.2484 | 4.67E-24 | 4.67E-23 |
| P53_PATHWAY | 200 | 42 | 0.21 | 3.85E-22 | 3.21E-21 |
| UV_RESPONSE_DNA_DAMAGE | 144 | 36 | 0.25 | 6.86E-22 | 4.90E-21 |
| GLYCOLYSIS | 200 | 41 | 0.205 | 3.13E-21 | 1.95E-20 |
| ESTROGEN_RESPONSE_EARLY | 200 | 40 | 0.2 | 2.46E-20 | 1.12E-19 |
| ESTROGEN_RESPONSE_LATE | 200 | 40 | 0.2 | 2.46E-20 | 1.12E-19 |
| MTORC1_SIGNALING | 200 | 40 | 0.2 | 2.46E-20 | 1.12E-19 |
| KRAS_SIGNALING_UP | 200 | 39 | 0.195 | 1.87E-19 | 7.20E-19 |
| MYOGENESIS | 200 | 39 | 0.195 | 1.87E-19 | 7.20E-19 |
| IL2_STAT5_SIGNALING | 200 | 37 | 0.185 | 9.86E-18 | 3.52E-17 |
| APICAL_JUNCTION | 200 | 33 | 0.165 | 1.82E-14 | 5.36E-14 |
| COMPLEMENT | 200 | 33 | 0.165 | 1.82E-14 | 5.36E-14 |
| INFLAMMATORY_RESPONSE | 200 | 33 | 0.165 | 1.82E-14 | 5.36E-14 |
| UV_RESPONSE_UPREGULATED | 158 | 29 | 0.1835 | 4.10E-14 | 1.14E-13 |
| INTERFERON_ALPHA_RESPONSE | 97 | 23 | 0.2371 | 5.81E-14 | 1.53E-13 |
| UNFOLDED_PROTEIN_RESPONSE | 113 | 24 | 0.2124 | 2.25E-13 | 5.63E-13 |

***Figure 51 - Over-represented pathways determined by the GSEA MSigDB database - ordered from the most significant p-value.***

Specific interacting pathways that the oriP-EBNA1 over-represented probe set interacted with as projected by the *MSigDB* GSEA software. GSEA provideed information on the number of probes which interact within each gene set (k) alongside p-value/FDR which relates to the specificity of the projections too. The processes are ordered from most significant p-value down.

The analysis using GSEA demonstrated, again, an interaction with a number of inflammatory and immune pathways such as *Interferon gamma response*, *Interferon alpha response* and *inflammatory response* (*Figure 51*). This corroborated and added greater validity to the inflammatory signature presented from the Reactome database.

Moreover, the most significantly projected pathway from the over-represented probe set was *Epithelial to Mesenchymal transition*. This represented unwanted cellular differentiation and suggested that a number of probes associated and over-represented in *oriP*-EBNA1 iPSCs were in relation to a loss of pluripotent phenotype.

Finally, the same analysis was undertaken using *Gene ontology (GO)* - a software of a similar ilk to *Reactome* and *GSEA*. GO would align the *oriP*-EBNA1 over-represented genes within their most specifically projected pathways. This would complete the unbiased analysis into key biological processes that *oriP*-EBNA1 iPSC-specific probes interacted within.

| Gene Set Name | Genes in Gene Set (K) | # Genes in Overlap (k) | k/K | p-value | FDR q-value |
|---|---|---|---|---|---|
| RESPONSE_TO_ENDOGENOUS_STIMULUS | 1648 | 198 | 0.1201 | 1.13E-58 | 8.33E-55 |
| REGULATION_OF_CELL_POPULATION_PROLIFERATION | 1708 | 199 | 0.1165 | 8.37E-57 | 3.08E-53 |
| BIOLOGICAL_ADHESION | 1417 | 177 | 0.1249 | 1.05E-54 | 2.57E-51 |
| CELLULAR_RESPONSE_TO_ENDOGENOUS_STIMULUS | 1384 | 171 | 0.1236 | 4.01E-52 | 7.37E-49 |
| NEGATIVE_REGULATION_OF_RESPONSE_TO_STIMULUS | 1655 | 185 | 0.1118 | 5.69E-50 | 8.37E-47 |
| CIRCULATORY_SYSTEM_DEVELOPMENT | 1161 | 152 | 0.1309 | 2.08E-49 | 2.55E-46 |
| RESPONSE_TO_GROWTH_FACTOR | 732 | 120 | 0.1639 | 3.38E-49 | 3.55E-46 |
| POSITIVE_REGULATION_OF_MULTICELLULAR_ORGANISMAL_PROCESS | 1795 | 191 | 0.1064 | 2.04E-48 | 1.87E-45 |
| APOPTOTIC_PROCESS | 1980 | 201 | 0.1015 | 7.16E-48 | 5.85E-45 |
| REGULATION_OF_CELL_DEATH | 1723 | 184 | 0.1068 | 8.37E-47 | 5.38E-44 |
| IOCOMOTION | 1943 | 197 | 0.1014 | 8.52E-47 | 5.38E-44 |
| EXTRACELLULAR_STRUCTURE_ORGANIZATION | 421 | 90 | 0.2138 | 8.79E-47 | 5.38E-44 |
| POSITIVE_REGULATION_OF_SIGNALING | 1828 | 189 | 0.1034 | 4.68E-46 | 2.65E-43 |
| CELL_MOTILITY | 1719 | 182 | 0.1059 | 9.54E-46 | 5.01E-43 |
| ANATOMICAL_STRUCTURE_FORMATION_INVOLVED_IN_MORPHOGENESIS | 1172 | 147 | 0.1254 | 1.81E-45 | 8.88E-43 |
| NEGATIVE_REGULATION_OF_SIGNALING | 1394 | 160 | 0.1148 | 1.62E-44 | 7.46E-42 |
| RESPONSE_TO_OXYGEN_CONTAINING_COMPOUND | 1616 | 171 | 0.1058 | 7.16E-43 | 3.10E-40 |
| REGULATION_OF_CELL_DIFFERENTIATION | 1863 | 185 | 0.0993 | 1.44E-42 | 5.89E-40 |
| CARDIOVASCULAR_SYSTEM_DEVELOPMENT | 806 | 117 | 0.1452 | 1.79E-42 | 6.92E-40 |
| RESPONSE_TO_CYTOKINE | 1192 | 143 | 0.12 | 6.07E-42 | 2.23E-39 |

***Figure 52 - Over-represented pathways determined by the GO database - ordered from the most significant p-value.***

Specific interacting pathways that the oriP-EBNA1 over-represented probe set interacts with as projected by the *GO* software. GO provides information on the number of probes which interact within each gene set (k) alongside p-value/FDR which relates to the specificity of the projections too. The processes are ordered from most significant p-value down.

Again, the analysis demonstrated an alignment with GO terms such as *Response to cytokines* and *Regulation of cell differentiation* as previously outlined using other gene enrichment software (*Figure 52).* However, GO terms associated with *Response to oxygen containing compound* were also significantly projected, bringing some context to previous terms in relation to *UV DNA damage response*.

Taken in its entirety, the initial probe enrichment analysis and biological process projection demonstrated that the over-represented probes in *oriP*-EBNA1-derived iPSCs were mostly associated with potentiating a pro-inflammatory cellular environment. Processes in relation to unwanted cellular differentiation were also commonalities between the analysis and helped to substantiate the spontaneous differentiation phenotype described in refractory lines which was negated when using the dbDNA system. Moreover, the analysis also demonstrated hints of a *DNA damage* phenotype alongside the presence of *oxygen containing compounds* in *oriP*-EBNA1 iPSCs.

Following this, several genes were isolated to validate their expression and over-representation within *oriP*-EBNA1 iPSCs. Surplus RNA that was taken at the same time for the microarray and was utilised to determine the expression of genes associated with some of the highlighted pathways within the probe enrichment analysis. Gene targets associated with the over-represented biological processes, signalling pathways and cellular processes projected by the *Reactome* database were isolated for analysis. An RT-qPCR analysis was undertaken to quantify the levels of gene expression within dbDNA and *oriP*-EBNA1 iPSC samples.

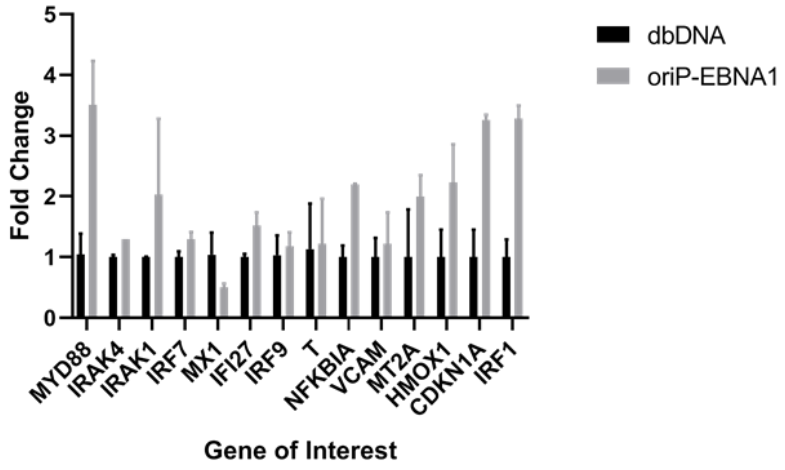



***Figure 53 - RT-qPCR analysis of EBNA1 over-represented probes using the Reactome database as a guide.***

The RT-qPCR represents a single experiment carried out on the remnants of RNA from cells which were isolated for the microarray experiment. Data represents n=1 with 3 technical repeats per gene. Error bars represent SEM for each gene.

The genes isolated for validation followed the same trend as expected with the microarray results, being largely over-expressed within *oriP*-EBNA1 iPSCs (*Figure 53*). This permits confidence in the accuracy of the projections determined by the enrichment factor software and guided us to dig further into the key transcriptomic differences and the potential mechanisms behind them.

#### 2.1.8.1. **STAT1 signalling and IFN-γ signalling:**

The gene enrichment analysis demonstrated that the genes over-represented within the *oriP*-EBNA1 produced iPSCs were largely associated with immune system processes. Moreover, upon analysing more specific, isolated biological processes it was clear that *interferon signalling* was a commonly over-represented within EBNA1-iPSCs. Using Reactome, we were able to substantiate the clear alignment of the over-represented EBNA probes to *interferon signalling* with a high degree of specificity, as demonstrated by the *p-value* (*Figure 54*).

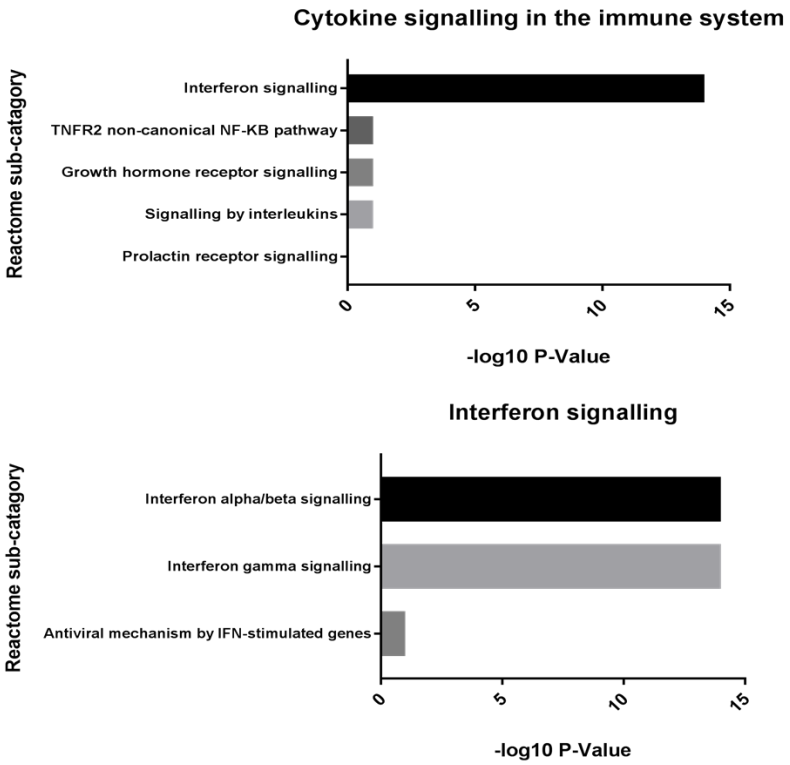



***Figure 54 – Reactome projection of oriP-EBNA1 over-represented probes in relation to interferon signaling.***

Graphs represent the projection of oriP-EBNA1 over-represented probes in relation to interferon signaling. The gprobes aligns with this biological process with a high specificity demonstrated by the p-value. The higher the absolute value of the statistic, the greater its significance and the greater the degree of specificity of the probe sets within this biological process.

To further substantiate the alignment of such *oriP*-EBNA1 over-represented genes within interferon related pathways, a heatmap analysis was generated. The analysis incorporated genes associated with interferon signalling pathways within both iPSC types and ESCs too. This would help to provide further context in relation to interferon-specific genes and if their expression was specifically over-represented within *oriP*-EBNA1 iPSCs or were they likewise shared by ESCs.

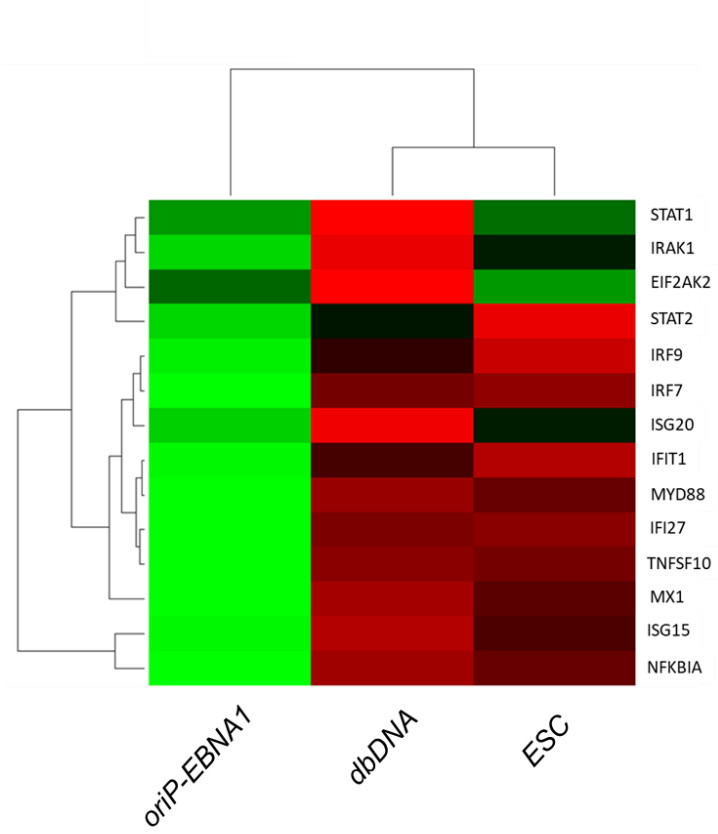



***Figure 55 - Heatmap representing genes associated with interferon expression in oriP-EBNA1 iPSCs, dbDNA iPSCs and ESCs.***

Heatmap analysis representing interferon-related gene expression within oriP-EBNA1 iPSCs, dbDNA iPSCs and ESCs too. The hierarchical clustering of the overall gene expression for this gene set demonstrates a close alignment of the expression pattern of both ESC and dbDNA iPSCs with oriP-EBNA1 iPSCs having the most divergent expression profile of such gene sets.

The heatmap of interferon-related genes demonstrated that they were largely over-represented in *oriP*-EBNA1 iPSCs specifically, suggesting that the vector may be responsible for the manipulation of such gene sets (*Figure 55*). The result was interesting and guided us to investigate the mechanism behind this further. Previous literature had demonstrated that the EBNA1 protein was associated with an increased activation of *STAT1* which in turn sensitised cells to an increase in *IFN-γ* expression and an activation of *Major Histocompatibility Complex* (MHC) (Wood et al., 2007). We therefore wanted to examine the effect of the *oriP*-EBNA1 vector and its ability to potentiate *STAT1* transcription in comparison to the dbDNA vector system. GFP-expressing *oriP*-EBNA1 and dbDNA vectors were transiently transfected into

HEK293T cells and every 24-hours over a 72-hour period the cells transcriptomics were analysed for *STAT1* expression using RT-qPCR.

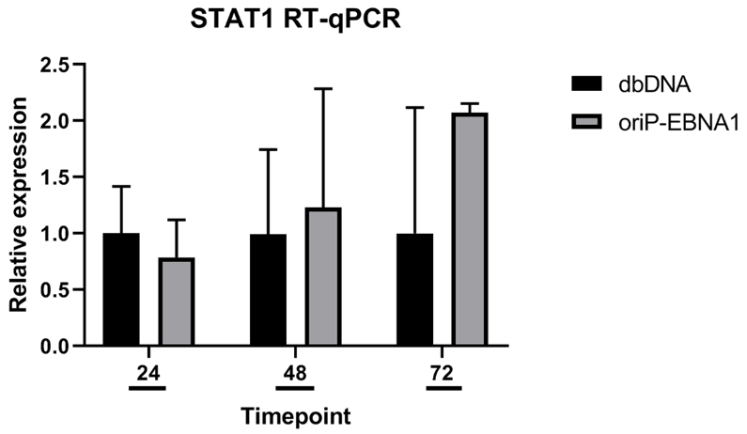



***Figure 56 - RT-qPCR quantifying the relative expression of STAT1 in cells transfected with both dbDNA and oriP-EBNA1 GFP expressing vectors.***

The graph represents an n=3 with 3 technical replicates per biological repeat. The data was subjected to a Kolmogorov-Smirnov test for normality with the p-value being >0.05 suggesting non-normal distribution. As such, a Mann-Whitney U-test was carried out per timepoint. * - p-value = <0.05, ** - p-value = <0.01, *** - p-value = <0.001. No significant difference was seen.

The RT-qPCR demonstrated an increased trend in *STAT1* transcription following transfection with the *oriP*-EBNA1 plasmid in comparison to the dbDNA system but not to a significant degree (*Figure 56*). Further experiments should be carried out to determine the subsequent localisation of *STAT1* expression following EBNA1 transfection to determine any differences in cell response and how this sensitises each transfected cell type to an increased propensity for interferon related signalling.

### 2.1.8.2. **Spontaneous differentiation of iPSCs:**

A commonality between the unbiased differentially expressed probe analysis carried out from the microarray was in relation to epithelial to mesenchymal transition/cellular differentiation. The dbDNA system had successfully reprogrammed primary fibroblast cultures with which the *oriP*-EBNA1 system could not produce stabilised iPSC lines from. Upon reprogramming such fibroblasts using the *oriP*-EBNA1 system, the iPSCs

would undergo spontaneous differentiation independent of any cues whilst the dbDNA produced iPSCs would maintain their pluripotent state indefinitely. Having compiled a list of markers of early differentiation, a comparison was carried out using expression values from the microarray for all the pluripotent cell types in a heatmap analysis.

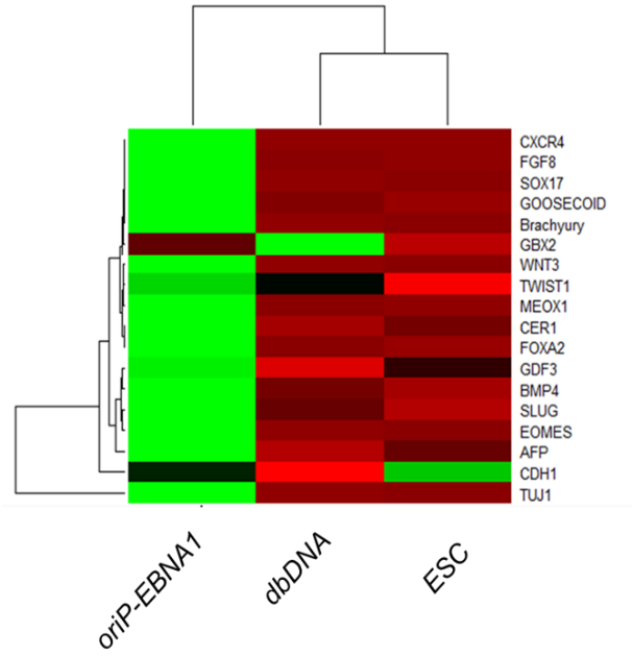



***Figure 57 - Heatmap analyzing the expression of key early markers for differentiation within oriP-EBNA1 iPSCs, dbDNA iPSCs and ESCs using data obtained from the microarray.***

Heatmap analysis representing markers of early differentiation within oriP-EBNA1 iPSCs, dbDNA iPSCs and ESCs. The hierarchical clustering of the overall probe expression for this gene set demonstrates a close alignment of the expression pattern of both ESC and dbDNA iPSCs with oriP-EBNA1 iPSCs having the most divergent expression profile of such probe sets.

The analysis indicates an over-representation of genes associated with early differentiation in iPSCs produced using the *oriP*-EBNA1 vector in relation to the dbDNA produced iPSCs and ESCs alike (*Figure 57*). Having evidenced such a pattern in the microarray, we then wanted to carry out a phenotypic analysis to determine the difference in the level of spontaneous differentiation between the two different iPSC types. The analysis was carried out by monitoring the level of spontaneous differentiation within the culture flasks for both dbDNA and *oriP*-EBNA1 produced iPSCs. 3 different primary fibroblast cultures were concurrently reprogrammed with both the dbDNA and *oriP*-EBNA1 vectors. The iPSCs produced were stained with AP to determine the area of spontaneous differentiation in proportion to the whole colony,

which was monitored and analysed using *imageJ*. The analysis was carried out on colonies from as early as p5 up to cells that were p32 in passage number. This was to account for any level instability amongst iPSCs at earlier iPSC passages.

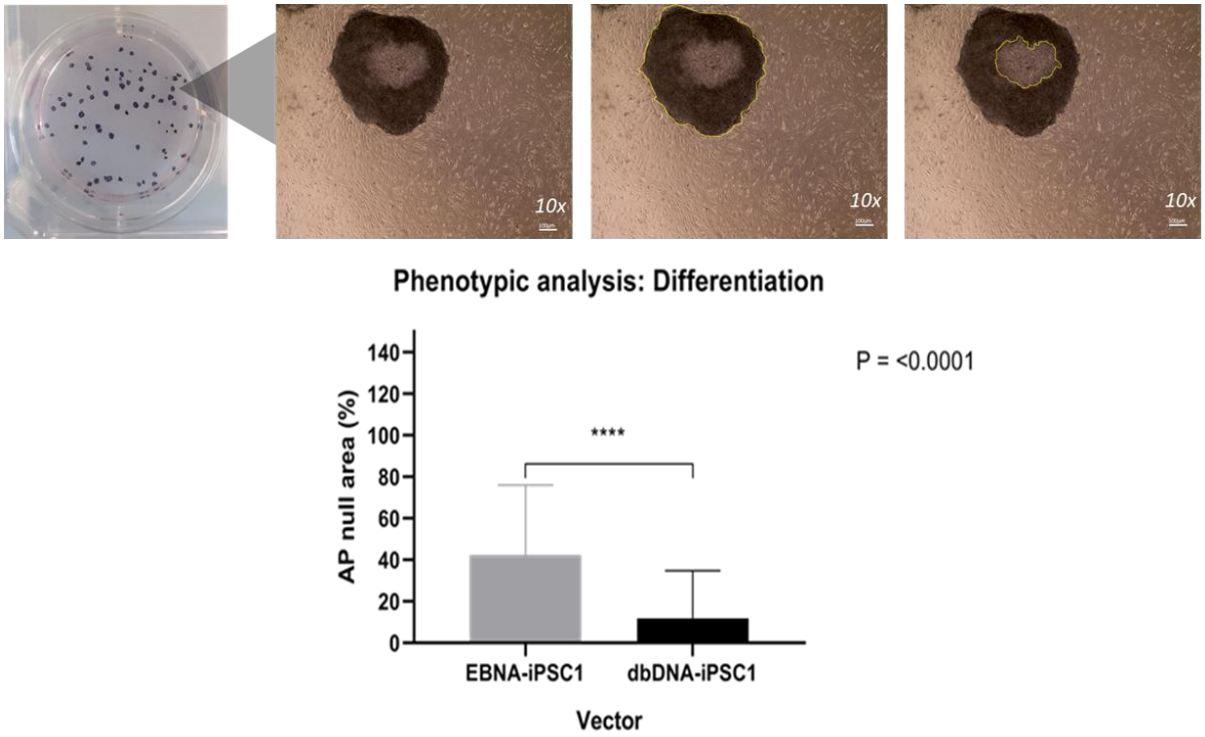



***Figure 58 – Analysis of spontaneous differentiation in both dbDNA and oriP-EBNA1 produced iPSCs.***

Representative images on how the level of phenotypic differentiation was quantified. Colonies underwent AP staining, which stains pluripotent areas and specifically excludes areas of spontaneous differentiation. The level of differentiation was then determined in proportion to the total area of the colony. This analysis was carried out across 3 different iPSC lines reprogrammed using both the oriP-EBNA1 and dbDNA systems. These iPSCs were concurrently reprogrammed and cultured simultaneously. The graph represents the average level of differentiation observed within iPSCs produced by both vector types. 3 different primary fibroblast cultures were analyzed with passages ranging from p5 to p32 with a minimum of 150 colonies being analyzed per iPSC line and vector. The data was subjected to a Kolmogorov-Smirnov test for normality with the p-value being >0.05 suggesting

The dataset evidenced unequivocally that the *oriP*-EBNA1 produced iPSC colonies displayed greater levels of spontaneous differentiation irrespective of the primary fibroblast cells used or the passage of the iPSCs analysed (*Figure 58*). The analysis provided a key quantification of the difference that was being observed routinely within the tissue culture environment in relation to the phenotypes of both iPSC cell types. The data clearly demonstrated, with a high significance, that the *oriP*-EBNA1 produced

iPSCs' consistently spontaneously differentiate at a higher rate than that of dbDNA produced iPSCs'.

#### 2.1.8.3. **DNA damage analysis:**

Previous literature had demonstrated a clear association between the EBNA1 protein, ROS production and a subsequent increased level of DNA damage (Gruhne et al., 2009). The dataset presented so far has demonstrated how EBNA1 upregulated probes were highly aligned with a number of biological processes associated with *response to oxygen containing compounds* and *DNA damage response*. Taken as a whole, there was clear rationale to investigate the potential of the *oriP*-EBNA1 system to induce DNA damage within transfected cells and how that relates in comparison to the dbDNA system. The comet assay was a well-established, simple method to detect DNA damage at a single cell level. The assay functioned on the premise that, when subjected to several buffers and electrophoresed, intact DNA will maintain its integrity and progress through the agarose gel as a unified entity. However, upon being damaged, this would result in smaller strands of DNA. When subjected to the same process, the smaller damaged DNA strands will run behind the intact DNA; creating a smear-like effect referred to as '*Tail DNA'*. *Tail DNA* was proportionate to the level of damaged DNA within a single cell – with the remaining *Head DNA* representing intact DNA. Software has also been developed with the capacity to quantify such damaged DNA per cell. Therefore, a comet assay was undertaken in order to determine the presence of any DNA damage within respective dbDNA and *oriP*-EBNA1 transfected cells. Aware however that incitements of DNA damage and the response to such damage may be minimal, we aimed to prime the cells to DNA damage induction by spiking the media with hydrogen peroxide to increase the sensitivity of the experiment. We therefore decided to titrate the level of hydrogen peroxide required to increase the sensitivity of the experiment whilst also to codetermine an appropriate concentration to utilise for a positive control.

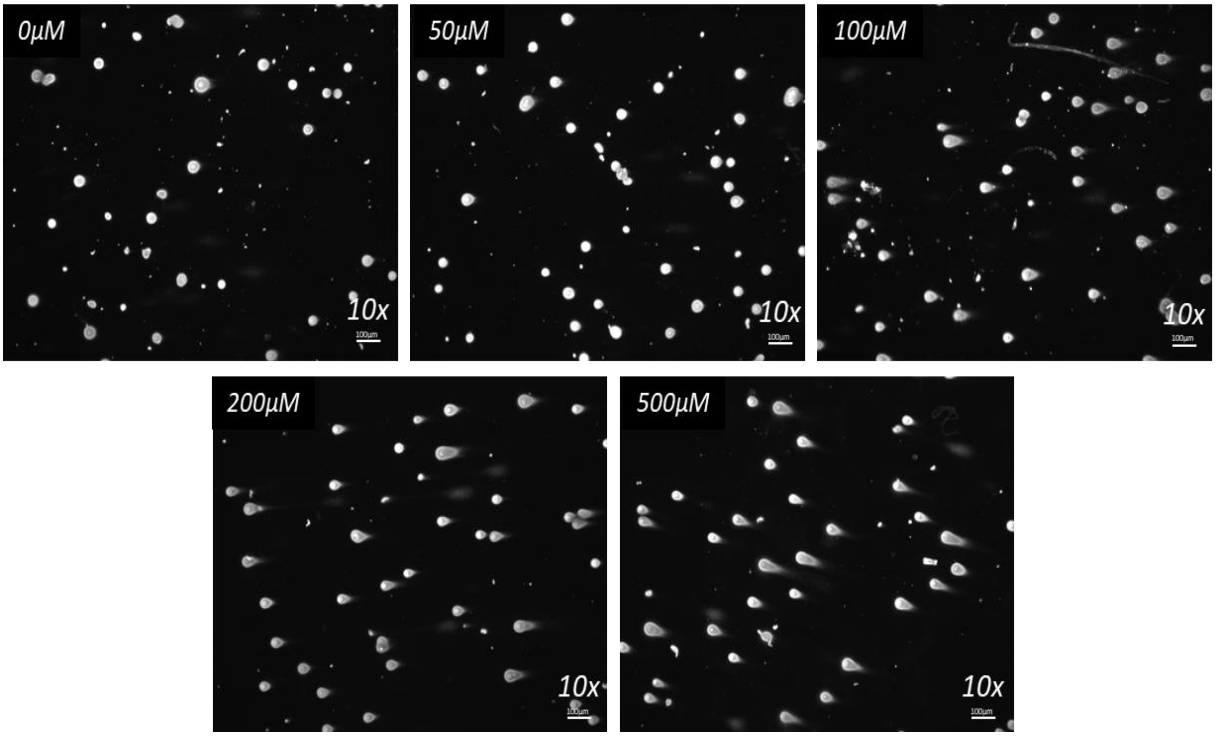



***Figure 59 - Respective comet assay images for the titration of different concentrations of hydrogen peroxide.***

Representative images of example comets for cells treated with different concentrations of hydrogen peroxide are depicted. The concentrations of hydrogen peroxide ranges from 0 to 500µM. Images were taken using a fluorescent microscope. Scale bar represents 100µm.

It was clear from the experiment that as the concentration of hydrogen peroxide increased the level DNA damage induction increased proportionately. This was indicated by the increased level of '*Tail DNA*' as the concentration of hydrogen peroxide the cells were treated with increased (*Figure 59*). To quantify *head DNA* and *tail DNA* a software called CaspLab was employed. Once thresholds, as detailed by the software developers were set, a measurement frame was drawn over cells to be analysed. The measurement frame was then be activated which develops a measurement profile calculating several parameters (*Head DNA*, *Tail DNA*, *Tail length, Head radius, Tail moment*). *Head DNA* accounts for the level of intact DNA and the *Tail DNA* accounts for DNA which was damaged within single cells. *Comet length* was also calculated which accounts for the length of the cell from the beginning of the *head* to the end of the *tail* and *Tail moment* which was a product of *tail length* and the percentage DNA in the *tail*. These measurements can then be exported for each

photo. This analysis was carried out for a minimum of 100 comets per gel with three repeats per concentration of hydrogen peroxide.

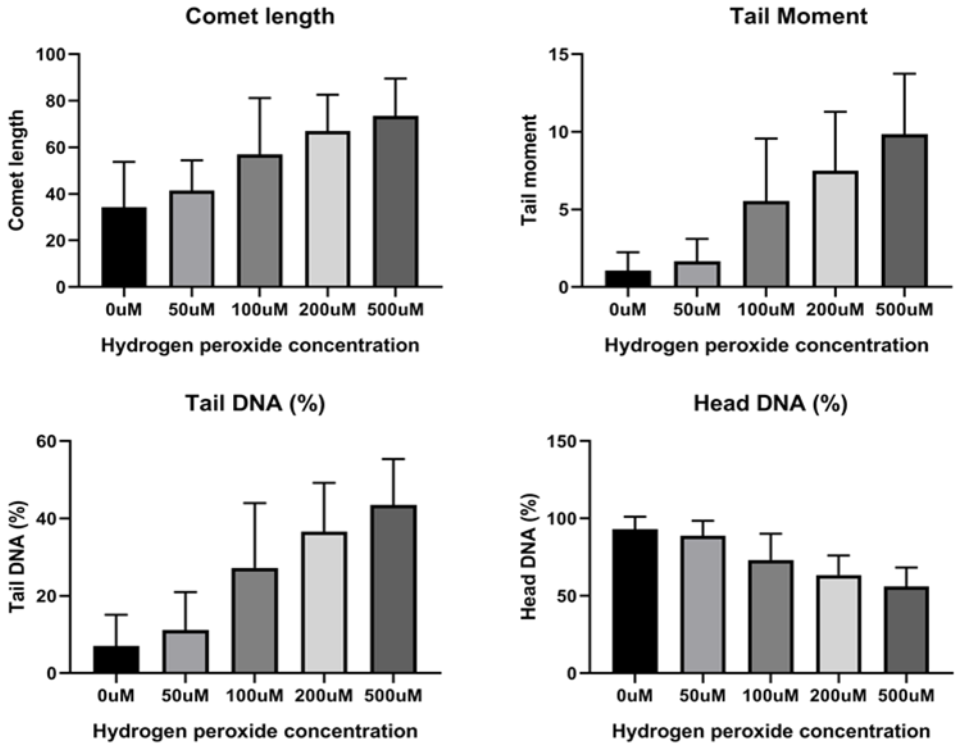



***Figure 60 - Titration of hydrogen peroxide levels added to HEK293T media before being analyzed using a comet assay.***

HEK293T cells were subjected to hydrogen peroxide for 30 minutes prior to analysis. CaspLab was then employed to quantify comet length, Tail moment, Tail DNA and Head DNA. 1 biological repeat with 3 technical repeats being carried out with a minimum of 100 cells per technical repeat being analysed. Error bars represent SEM for each biological repeat.

The results demonstrated a proportionate increase in DNA damage as the concentration of hydrogen peroxide increased (*Figure 60*). This was demonstrated in several parameters that were calculated using CaspLab. *Head DNA* (%) decreased as the concentration of hydrogen peroxide increased. As *head DNA* was proportionate to intact DNA there will also be a correlation between the level of *head DNA* and the level of *tail DNA* which was captured. As *Head DNA* decreased this will be secondary to damaged DNA and therefore an increased level of *tail DNA*. *Tail DNA* proportionately increased as the concentration of hydrogen peroxide increased demonstrating a greater level of DNA damage. Again, *comet length* was also related

to the level of *tail DNA*. When a cell exhibits more damaged DNA, the *tail* produced should in theory be longer and therefore the cells should display a greater *comet length*. This was corroborated by our results as the *comet length* likewise proportionately increased along with the concentration of hydrogen peroxide and DNA damage. Finally*, tail moment*, proportionate to *tail length* and DNA within the *tail* likewise increased proportionately with an increase in hydrogen peroxide secondary to a higher level of DNA damage. It was clear from our data that a phenotype of DNA damage was more apparent in the comet assay as the concentration of hydrogen peroxide increased. The results suggested that 50µM initiated small incitements of DNA damage that was detectable by the comet assay but produced results that were not largely dissimilar to untreated cells. This was therefore deemed an appropriate concentration of hydrogen peroxide to prime all cells to be analysed with to increase the sensitivity of the assay. A concentration of 200µM of hydrogen peroxide induced incitements of DNA damage that were much more extreme in comparison to untreated cells and was therefore utilised as a positive control concentration for the assay. From this, an experiment comparing and quantifying DNA damage induction in cells transfected with the dbDNA and *oriP*-EBNA1 vectors was carried out. HEK293T cells were transiently transfected with 4µg of dbDNA-eGFP ($6.12x10^{35}$) and 4µg *oriP*-EBNA1-eGFP ($3.4x10^{35}$). The cells were left for 24 hours before the level of GFP-expression was qualified.

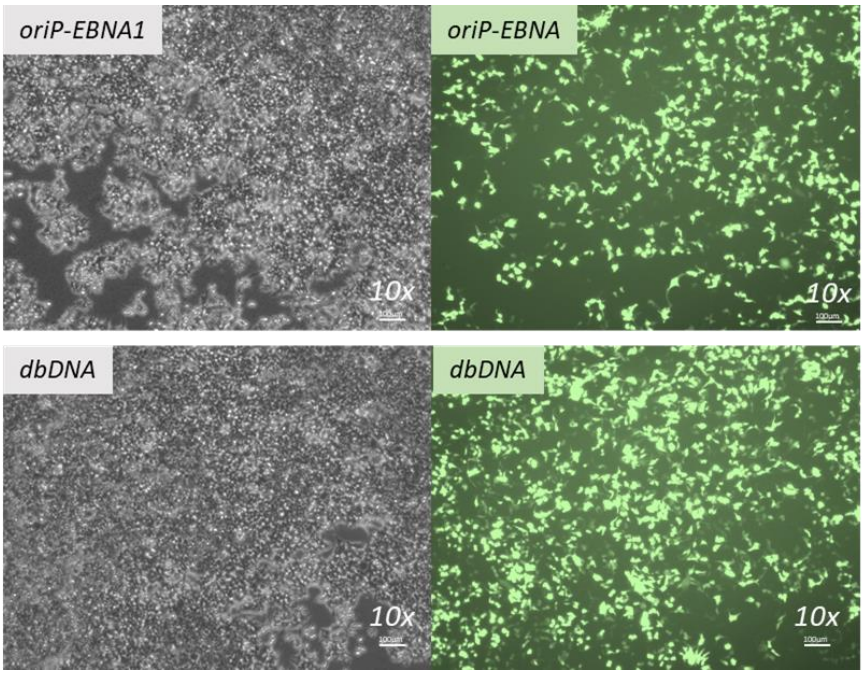



***Figure 61 - eGFP qualification of both the oriP-EBNA1 and dbDNA vector following transfection into HEK293T cells.***

HEK293T cells were transfected with dbDNA and oriP-EBNA1 vectors both with an added GFP transgene to determine transfection efficiency and success. The GFP qualification demonstrates the transfection was successful and that the cells have the oriP-EBNA1 and dbDNA vectors within them. Qualification was undertaken 24 hours post PEI transfection.

GFP qualification demonstrated that a good transfection efficiency for both the dbDNA or *oriP*-EBNA1 vectors (*Figure 61*). The cells were then spiked with 50µM of hydrogen peroxide for 30 minutes, with positive controls being spiked with 200µM for the same time period. The cells were then subjected to a comet assay and analysed using the CaspLab software.

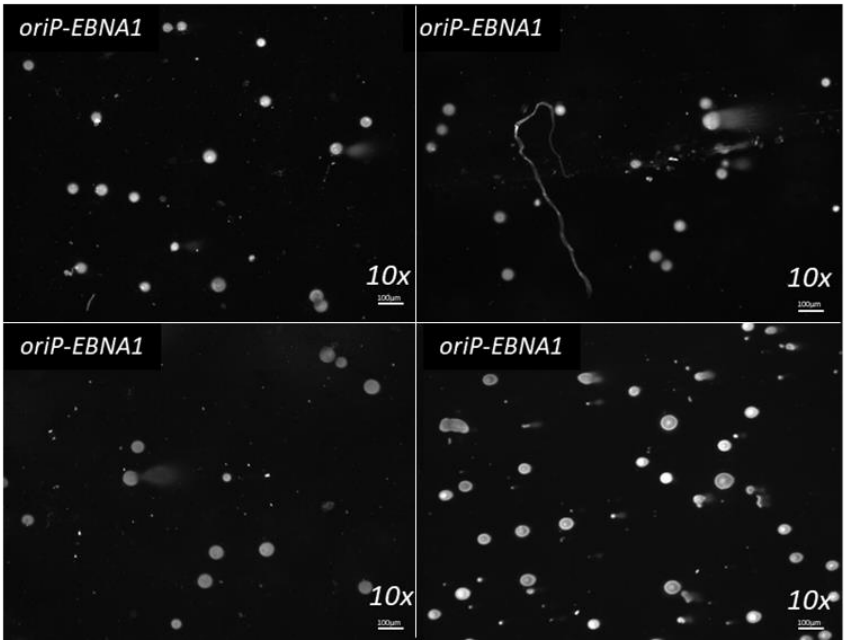



***Figure 62 - Example representative images of HEK293T cells following transient transfection with the oriP-EBNA1 eGFP vector.***

HEK293T cells which have been transfected with the oriP-EBNA1 GFP vector were processed in the comet assay with representative images being taken for analysis using CaspLab. An n=3 was taken with 3 technical repeats per biological replicate. 100 cells were images per gel. Error bar represents 100um.

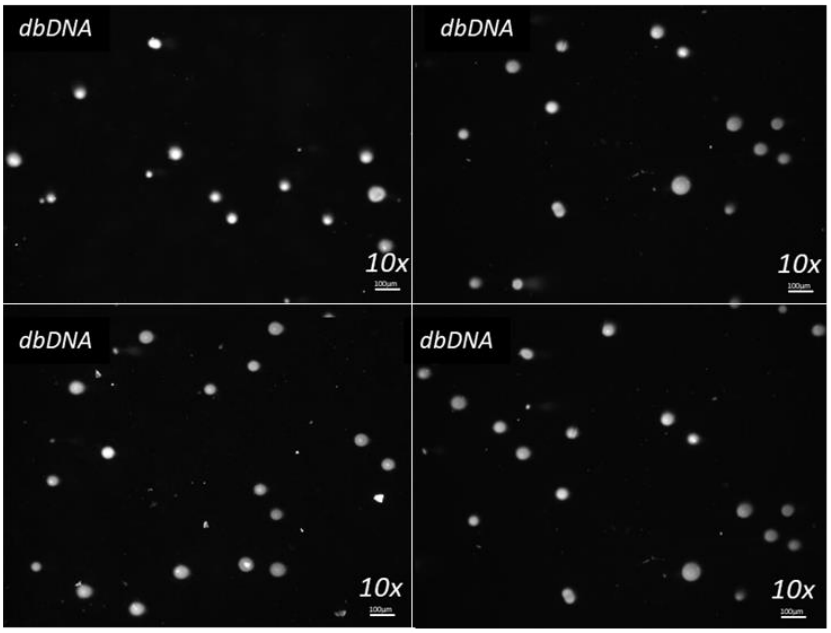



***Figure 63 - Example representative images of HEK293T cells following transient transfection with the dbDNA- eGFP vector.***

HEK293T cells which have been transfected with the dbDNA GFP vector were processed in the comet assay with representative images being taken for analysis using CaspLab. An n=3 was taken with 3 technical repeats per biological replicate. 100 cells were images per gel. Error bar represents 100um.

Representative images for both vector types were taken and stored before being analysed using the CaspLab software (*Figure 62 & Figure 63).*

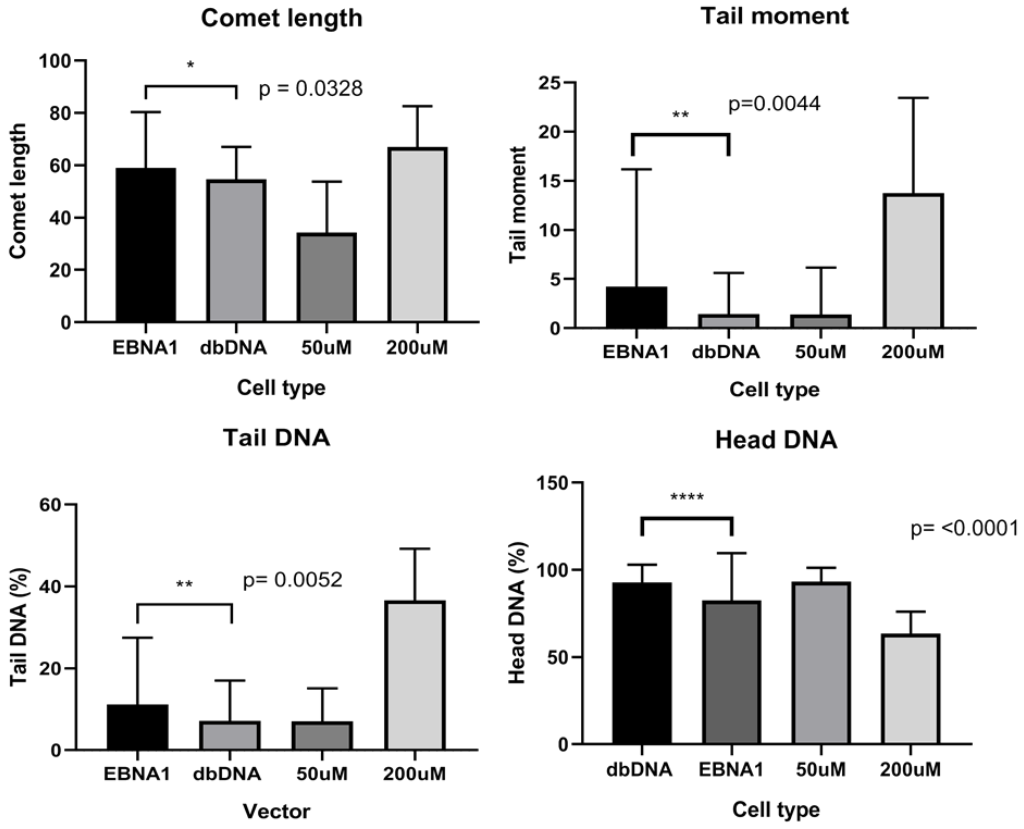



***Figure 64 – Analysis of HEK293T cells transfected with dbDNA and oriP-EBNA1 vectors using CaspLab software.***

Graphs depicting CaspLab generated results examining comet length, tail moment, tail DNA and head DNA. Each graph represents an n=3 experiment with 3 technical repeats per biological repeat. A minimum of 100 comets was analyzed per technical repeat. Each data type was subjected to a Kolmogorov-Smirnov test for normality with the p-values being >0.05 suggesting non-normal distribution for all graphs. As such, a Mann-Whitney U-test was carried out between dbDNA and oriP-EBNA1 transfected cells. * - p-value = <0.05, ** - p-value = <0.01, *** - p-value = <0.001, **** -p-value = <0.0001.

The results in demonstrated that *oriP*-EBNA1 transfected cells displayed a greater level of DNA damage in comparison to dbDNA transfected cells to a significant degree (*Figure 64*). *oriP*-EBNA1 transfected cells demonstrated greater levels of *tail DNA* and *tail moment* alongside a greater *comet length* than dbDNA transfected cells. Moreover, the level of intact or *head DNA* was significantly higher in dbDNA transfected cells; corroborating that the *oriP*-EBNA1 vector resulted in a greater induction/reduced repair of damaged DNA within transfected cells. To further substantiate this dataset, supporting experiments were undertaken.

An analysis incorporating the use of a γ-H2AX antibody was determined to be applicable for further experiments to corroborate differences in DNA damage incitement between the two vector systems. The γ-H2AX antibody localises to variant H2A histones which, upon the incitement of DNA damage, replace conventional H2A subsets and become rapidly phosphorylated. As such, a γ-H2AX antibody can easily detect and manifest discrete nuclear foci which can be utilised to enumerate the volume of DNA double strand breaks (DSB) in a cell. We nucleofected fibroblasts with both *oriP*-EBNA1 and dbDNA GFP-expressing vectors to determine if the transfection of either of these vector systems resulted in an increased level of variant/γ-H2AX. The use of the GFP vectors would allow for a more specific analysis too, as when carrying out ICC the GFP protein was retained allowing us to pinpoint and enumerate γ-H2AX in transfected cells specifically. Again, hydrogen peroxide was incorporated to induce DNA damage for a positive control too.

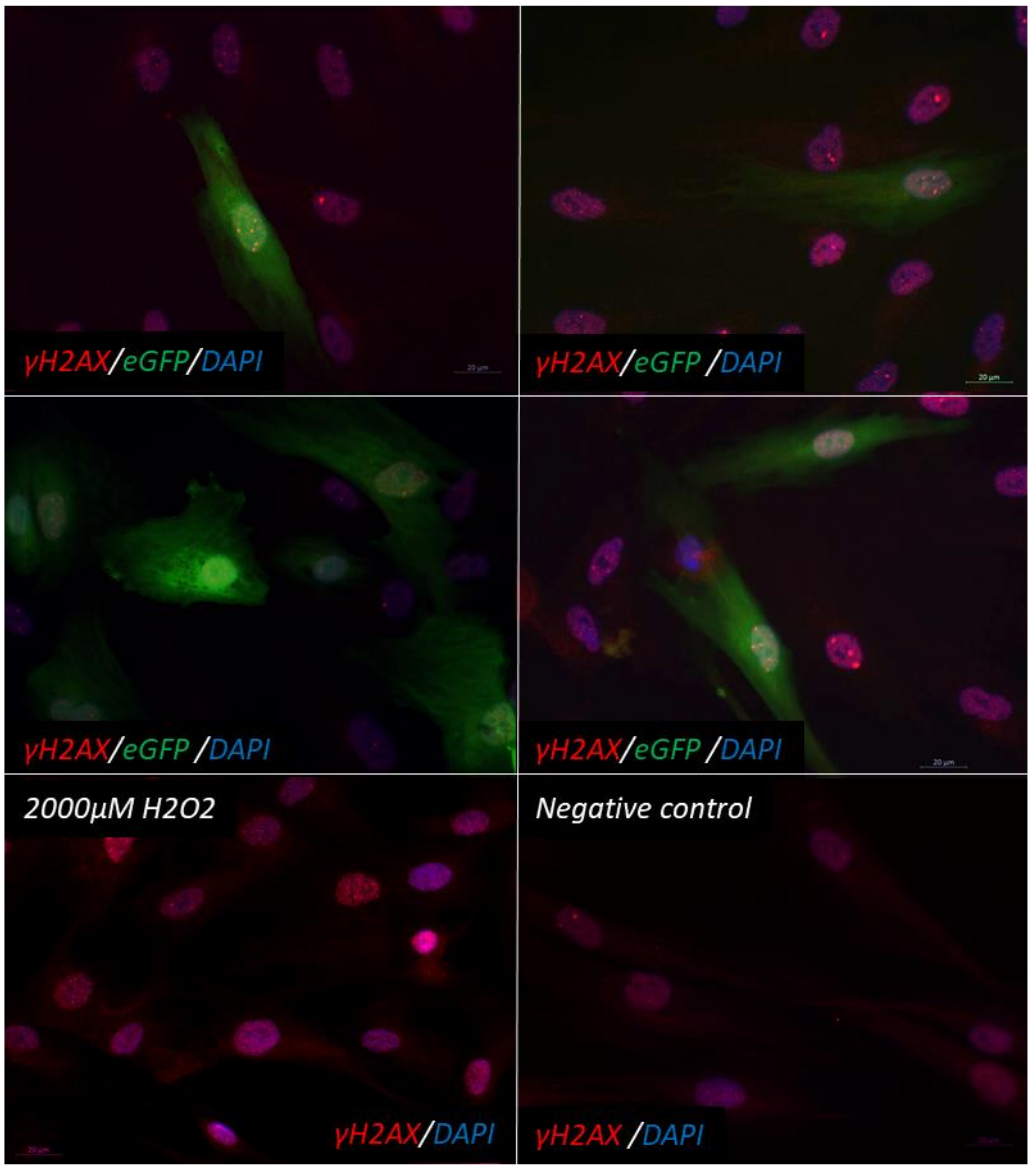



***Figure 65 - A γ-H2AX antibody was utilized for ICC being carried out on dbDNA/oriP-EBNA1 eGFP transfected cells.***

A γ-H2AX antibody was utilized to stain cells to determine the level of variant H2A within cells in proportion to the level of DSB. Both dbDNA and oriP-EBNA1 GFP expressing vectors were transfected into fibroblasts before being analyzed 24 hours after transfection. Fluorescent microscopy was utilized to determine GFP expressing cells which were then analyzed for γ-H2AX foci. A positive control utilizing 2000uM hydrogen peroxide and a negative control was also utilized. An n=3 was analyzed with 3 technical repeats per biological replicate. A minimum of 100 transfected/green cells were analyzed per technical repeat.

The immunostain demonstrated foci being present in all conditions of the experiment (*Figure 65*). The foci representing the level of DSB was then quantified for each cell type.

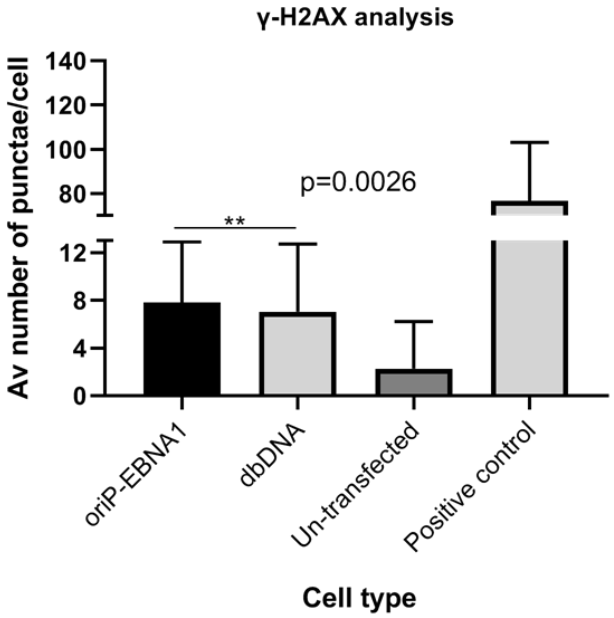



***Figure 66 - Quantification of the number of punctae within transfected cells following ICC staining using a γ-H2AX antibody.***

Representative images taken following ICC analysis using a γ-H2AX antibody. An n=3 was carried out for this experiment with 3 technical repeats per biological replicate. A minimum of 150 transfected cells per biological replicate were analysed. The data was subjected to a Kolmogorov-Smirnov test for normality with the p-value being >0.05 suggesting non-normal distribution of the dataset. As such, a Mann-Whitney U-test was carried out between dbDNA and oriP-EBNA1 transfected cells. * - p-value = <0.05, ** - p-value = <0.01, *** - p-value = <0.001.

Following quantification, it was clear that there was consistently a greater level of γ-H2AX foci in *oriP*-EBNA1 transfected cell types in comparison to dbDNA transfected cells (*Figure 66*). This demonstrated that there was a greater level of DSB present in *oriP*-EBNA1 transfected cells.

Altogether, the comet assay and γ-H2AX results demonstrated that *oriP*-EBNA1 transfected cells display an increased DNA damage phenotype in comparison to dbDNA transfected cells. This moreover corroborates some of the phenotypes denoted from the microarray analysis alongside what has been noted in previous literature too. Potentiating and inciting a reduced level of DNA damage was a real benefit to the use of the dbDNA system within reprogramming and iPSC development, as genomic integrity is an essential consideration for the downstream application of such pluripotent cell types.

### 2.1.9. **Over-represented genes within dbDNA-iPSCs:**

For a well-rounded interrogation of the microarray dataset, genes which were over-represented within the dbDNA produced iPSCs were also analysed. As per the *oriP*-EBNA1 system, the aim was to investigate what probes was significantly upregulated in dbDNA iPSCs in comparison to *oriP*-EBNA1 produced iPSCs and what biological processes such probes interact within. Over-represented dbDNA probes was compiled into a heatmap analysis with *oriP*-EBNA1-iPSCs and ESCs. The aim, as before, was to determine if the over-represented probes within the dbDNA system were specific to that iPSC-type only or, if the expression profile of this gene set was more like that of ESCs. The latter would most likely make it an insufficiency on behalf of the *oriP*-EBNA1 system to therefore upregulate said probes.

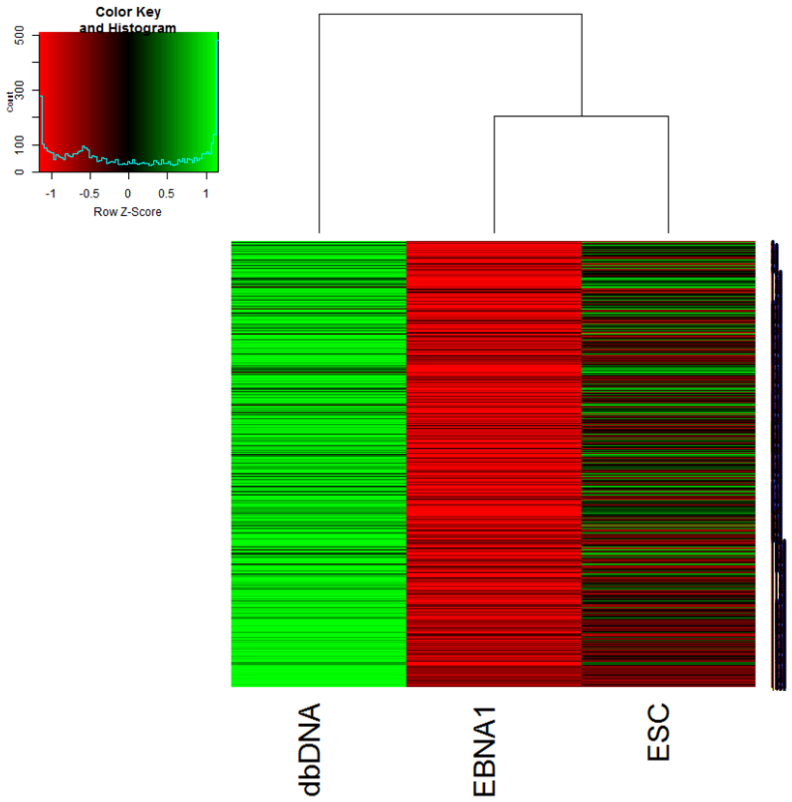



***Figure 67 - Heatmap comparison of probes that are over-represented in dbDNA iPSCs in comparison with oriP-EBNA1 iPSCs and ESCs.***

A heatmap was produced using R-studio. Image depicts all the over-represented gene probes within dbDNA iPSCs with green being highly expressed, black in-between and red a low level of expression. Hierarchical clustering also utilized to determine the similarities between cell types in respect to the gene sets.

The heatmap analysis suggests that the probes which were differentially expressed by dbDNA produced iPSCs were mostly unique to this cell type. Both *oriP*-EBNA1 iPSCs and ESCs share a more similar expression pattern in relation to the same genes, suggesting a degree of manipulation in a dbDNA dependent and pluripotency independent fashion (*Figure 67*). The probes which were significantly over-represented within the dbDNA system were then isolated for analysis to determine the biological processes within which such over-represented genes interact within. The *Reactome* database was utilised to project significant genes onto the human genome to help elucidate their interacting pathways.

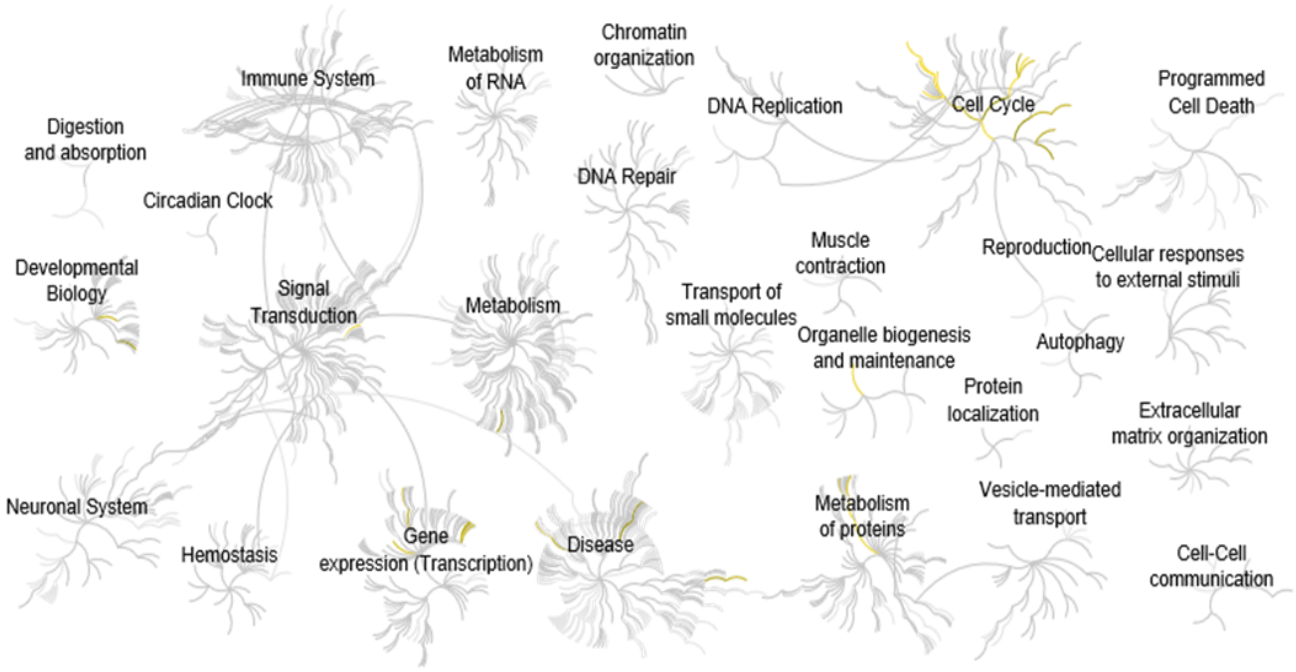



***Figure 68 - Projection of dbDNA over-represented probes onto the human genome as determined by the Reactome database.***

dbDNA over-represented probes were incorporated into the Reactome database which projects them onto different biological processes with which the genes most specifically align with. Each pathway has a focal point in the center with relating pathways being represented as branches protruding from this center point. Any pathways determined to be significantly interacting with the upregulated genes input into the system will be highlighted as yellow. FDR cut off <0.05 was employed.

Reactome projections aligned most of the dbDNA over-represented probes within the cell cycle biological processes with a high degree of specificity (*Figure 68*). Having previously detailed that the dbDNA system reprograms fibroblasts back to a pluripotent state independent of the *shp53* transgene, it was therefore possible that the exclusion of this protein may result in cell cycling differences between these two cell types. However, in order to gain greater insight, a detailed analysis looking at individual pathways was carried out. To remove any bias from the analysis, biological processes with which dbDNA over-represented probes interacted within were again determined from 3 separate sources: *Reactome, Gene set enrichment analysis* (*GSEA*) and *Gene ontology* (*GO*).

| Gene Set Name | Genes in Gene Set (K) | #Genes in Overlap (k) | k/K | p-value | FDR q-value |
|---|---|---|---|---|---|
| E2F_TARGETS | 200 | 41 | 0.205 | 2.86E-21 | 1.43E-19 |
| G2M_CHECKPOINT | 200 | 31 | 0.155 | 5.95E-13 | 1.49E-11 |
| MITOTIC_SPINDLE | 199 | 28 | 0.1407 | 8.16E-11 | 1.36E-09 |
| GLYCOLYSIS | 200 | 23 | 0.115 | 1.77E-07 | 2.22E-06 |
| UV_RESPONSE_DN | 144 | 19 | 0.1319 | 2.41E-07 | 2.41E-06 |
| ADIPOGENESIS | 200 | 22 | 0.11 | 7.02E-07 | 5.01E-06 |
| HEME_METABOLISM | 200 | 22 | 0.11 | 7.02E-07 | 5.01E-06 |
| DNA_REPAIR | 150 | 18 | 0.12 | 2.02E-06 | 1.27E-05 |
| APICAL_JUNCTION | 200 | 21 | 0.105 | 2.64E-06 | 1.32E-05 |
| MYC_TARGETS_V1 | 200 | 21 | 0.105 | 2.64E-06 | 1.32E-05 |

***Figure 69 - Over-represented pathways determined by the Reactome database - ordered from the most significant p-value.***

Specific interacting pathways that dbDNA over-represented probes interacted with as projected by the Reactome database. Reactome provides information on the number of entities which interact within this process and likewise the entities p-value/FDR which relates to the specificity of the projections too. The processes were ordered from most significant p-value down.

The specific biological processes as outlined by *Reactome* demonstrated again that a lot of the over-represented genes were interacting within cell cycle processes (*Figure 69*). *E2F targets* could embody several different genes and transcription factors which have both progressive and repressive cell cycle functions, but whose functionality was intrinsically linked to cell cycling. *G2M checkpoint* and *mitotic spindle* again were biological processes which were linked to efficient cell cycle functionality. Moreover, DNA repair was also highlighted which may again be manipulated secondary to the lack of *shp53* presence within the dbDNA reprogramming vector. The same analysis was subsequently undertaken using GSEA likewise.

| Gene Set Name | Genes in Gene Set (K) | Genes in Overlap (k) | k/K | p-value | FDR q-value |
|---|---|---|---|---|---|
| CELL_CYCLE | 642 | 80 | 0.1246 | 7.48E-25 | 1.12E-21 |
| CELL_CYCLE_MITOTIC | 536 | 68 | 0.1269 | 9.95E-22 | 7.46E-19 |
| POST_TRANSLATIONAL_PROTEIN_MODIFICATION | 1429 | 110 | 0.077 | 6.56E-17 | 3.28E-14 |
| MITOTIC_PROMETAPHASE | 198 | 35 | 0.1768 | 3.06E-16 | 1.15E-13 |
| GENE_EXPRESSION_TRANSCRIPTION | 1486 | 108 | 0.0727 | 6.27E-15 | 1.88E-12 |
| M_PHASE | 392 | 46 | 0.1173 | 7.30E-14 | 1.82E-11 |
| METABOLISM_OF_LIPIDS | 738 | 66 | 0.0894 | 1.78E-13 | 3.81E-11 |
| DEVELOPMENTAL_BIOLOGY | 1104 | 85 | 0.077 | 2.62E-13 | 4.90E-11 |
| AXON_GUIDANCE | 554 | 53 | 0.0957 | 3.46E-12 | 5.76E-10 |
| REGULATION_OF_PLK1_ACTIVITY_AT_G2_M_TRANSITION | 87 | 20 | 0.2299 | 4.48E-12 | 6.71E-10 |

***Figure 70 - Over-represented pathways determined by the GSEA MSigDB database - ordered from the most significant p-value.***

Specific interacting pathways that the dbDNA over-represented probes interacted with as projected by the MSigDB GSEA software. GSEA provides information on the number of genes which interact within each gene set (k) alongside p-value/FDR which relates to the specificity of the projections too. The processes were ordered from most significant p-value down*.*

Again, the GSEA software highlighted a number of biological processes associated with cell cycling as being specifically associated with the dbDNA over-represented probes (*Figure 70*). Five of the top 10 biological processes identified using GSEA were in relation to cell cycle regulation and progression, a common factor identified between the first two analysis sets. Finally, the Gene ontology (GO) database was also utilised for the same type of analysis.

| Gene Set Name | Genes in Gene Set (K) | Genes in Overlap (k) | k/K | p-value | FDR q-value |
|---|---|---|---|---|---|
| CELL_CYCLE | 1847 | 191 | 0.1034 | 9.86E-47 | 7.25E-43 |
| CELL_CYCLE_PROCESS | 1383 | 150 | 0.1085 | 6.82E-39 | 2.51E-35 |
| MITOTIC_CELL_CYCLE | 1009 | 121 | 0.1199 | 1.43E-35 | 3.50E-32 |
| NEUROGENESIS | 1599 | 148 | 0.0926 | 1.11E-30 | 1.89E-27 |
| CELL_PROJECTION_ORGANIZATION | 1512 | 143 | 0.0946 | 1.28E-30 | 1.89E-27 |
| CYTOSKELETON_ORGANIZATION | 1298 | 127 | 0.0978 | 1.37E-28 | 1.68E-25 |
| SMALL_MOLECULE_METABOLIC_PROCESS | 1688 | 146 | 0.0865 | 3.28E-27 | 3.44E-24 |
| NEURON_DIFFERENTIATION | 1348 | 127 | 0.0942 | 4.66E-27 | 4.28E-24 |
| NEGATIVE_REGULATION_OF_RNA_BIOSYNTHETIC_PROCESS | 1246 | 121 | 0.0971 | 5.76E-27 | 4.71E-24 |
| PROTEIN_CONTAINING_COMPLEX_ASSEMBLY | 1911 | 157 | 0.0822 | 7.49E-27 | 5.50E-24 |

***Figure 71 - Over-represented pathways determined by the GO database - ordered from the most significant p-value.***

Specific interacting pathways that the oriP-EBNA1 over-represented probes interacted with as projected by the *GO* software. GO provides information on the number of genes which interact within each gene set (k) alongside p-value/FDR which relates to the specificity of the projections too. The processes were ordered from most significant p-value down.

As anticipated, the analysis again yielded results with cell cycle functions being the most significantly aligned (*Figure 71*).

#### 2.1.9.1. **Cell cycle analysis:**

Following the unbiased probe/gene enrichment and biological process analysis, it was clear that differences involving genes associated with cell cycling were a commonality amongst the results. The results indicated that differences may exist between cell cycle checkpoints between the two iPS cell types (potentially in relation to *shp53* presence) moreover, *G2/M transition* was also highlighted as a commonality within the dataset. Subsequently, a cell cycle analysis between dbDNA and *oriP*-EBNA1 iPSCs was carried out to determine any differences between the two cell types.

The analysis was carried out using a Propidium Iodide (PI) stain before analysing the cells using a flow cytometer. PI was a stain that would intercalate between bases within nuclear DNA with little or no preference. Then, dependent on the cycling phase exhibited by a cell, this will dictate the amount of stain taken up and subsequently the level of fluorescence presented by a cell. In a crude simplification, cells will often go through an initial growth phase in G1, before entering the S phase. During the S phase, cells will undergo DNA replication before progressing into the G2 phase. Cells within the G2 phase will then prepare to divide until ultimately you have cell division or mitosis. When incorporated into this explanation, it was clear that cells in the G1 phase will have the least amount of DNA and as such a proportionately limited intercalation of the PI stain. Cells in the S phase however, will have more DNA than when they are in the G1 phase and will continue to fluoresce more brightly until the DNA has doubled its content. Therefore, cells in the G2 phase will incorporate twice as much PI as cells in G1 phase and therefore fluoresce with a greater intensity. It was therefore possible to determine, using this stain, differences between the cell cycle phases exhibited between dbDNA and *oriP*-EBNA1 produced iPSCs. This experiment was carried out on dbDNA and *oriP*-EBNA1 produced iPSCs using PI. Positive stain detection was carried out using flow cytometry.

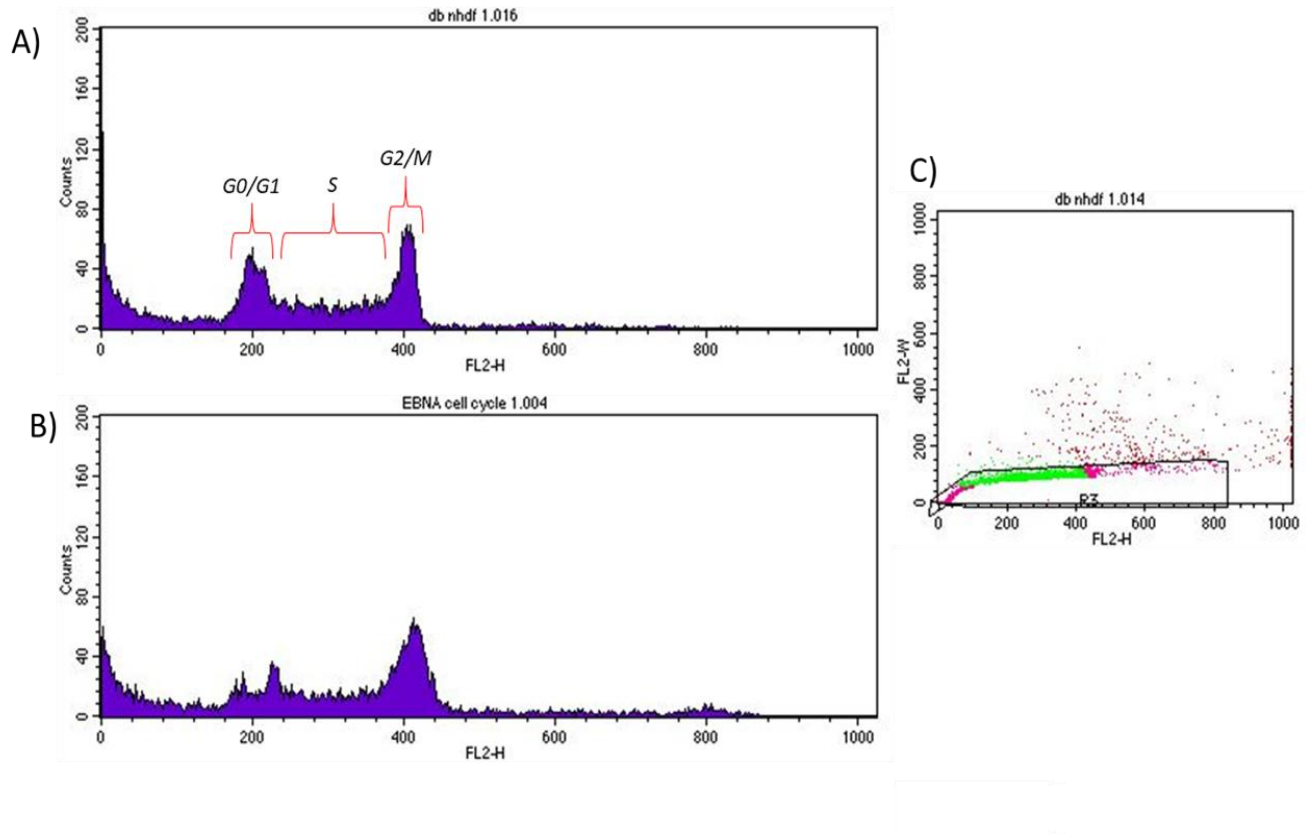



***Figure 72 - Representative histogram generated from both dbDNA and oriP-EBNA1 produced iPSCs following staining with PI any analysis using a flow cytometer.***

Representative histograms generated by the *BD FACS Caliber* upon subsequent analysis of iPSCs following staining with PI. A) represents dbDNA produced iPSCs stained with PI following flow cytometry analysis. The 5 different cell cycle phases are demonstrated amongst the 3 peaks beginning from ~200 on the FL2-H x-axis. This first peak depicts G0/G1, with the in-between region (between 200-400 FL2-H) demonstrating the S-Phase before the final peak at ~400 on the x-axis representing G2/M. B) represents the same analysis utilizing oriP-EBNA1 produced iPSCs. C) represents extra gating undertaken to remove doublets from the analysis, ensuring there was minimal contamination of doublets being recognized within the G2/M phase peak.

The results produced from the stained iPSCs from both cell types demonstrated that different phases were favoured more commonly between the two iPSC-types. The representative histogram (‘A)’) demonstrate that dbDNA-produced iPSCs seemingly had a greater proportion of its iPSCs within the *G0/G1* phases upon analysis in comparison to the *oriP*-EBNA1 (‘B)’) histogram. While, the oriP-EBNA1 iPS cells seemingly have more cells in the *G2/M* phase of cell cycle (*Figure 72*). However, in order to truly elucidate and quantify differences between the two cell types, the number of cells within each phase needed to be quantified. The resulting data was then

subsequently applied to *ModFit LT 5.0* permitting the quantification of the different cell cycle phases demonstrated within each cell type.

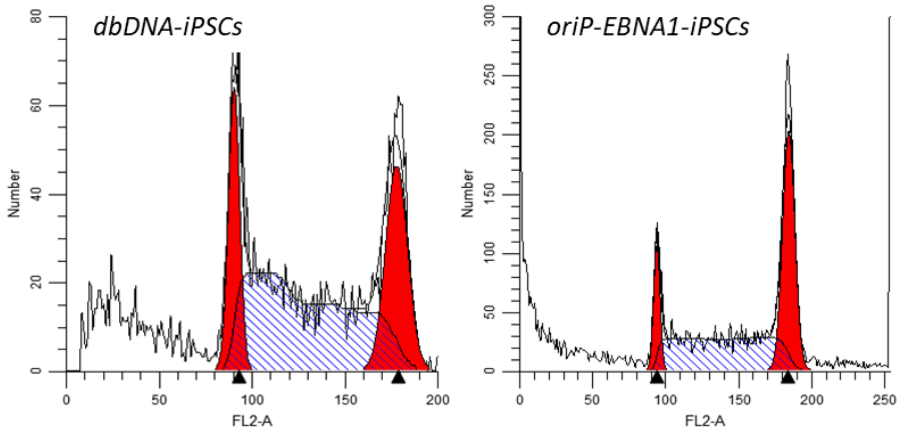



***Figure 73 - Example ModFit analysis of representative histograms for both dbDNA and oriP-EBNA1 produced iPSCs.***

Following cell cycle analysis, the subsequent histograms can be applied to the ModFit program which can then quantify the proportion of a cell population within each different phase for both cell types. The analysis was carried out for both dbDNA and oriP-EBNA1 produced iPSCs. An n=3 was analyzed with 3 technical repeats per biological replicate.

The analysis using ModFit was able to provide a quantification of the proportion of cells within each phase for both *oriP*-EBNA1 and dbDNA produced iPSCs (*Figure 73*). The analysis was undertaken for 3 different iPSC lines for each vector type produced from 3 different fibroblast sources. This was then collated and placed in a bar chart to examine the differences between the proportion of cells within each cell cycle phase for both iPSC types.

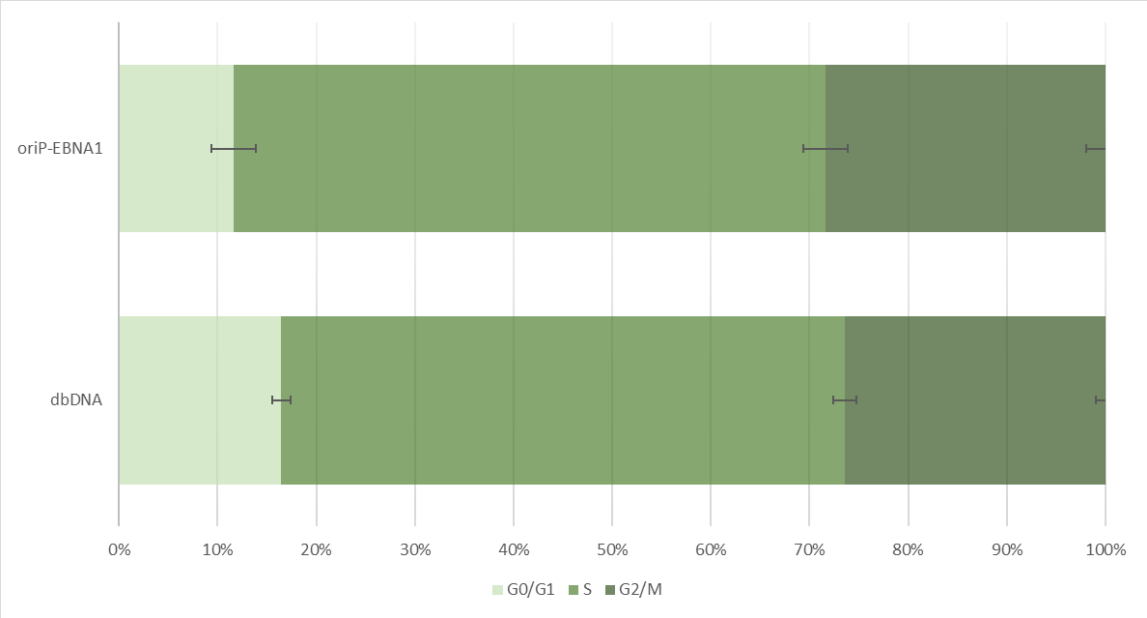



***Figure 74 - Graph representing the average cell cycle of both oriP-EBNA1 and dbDNA iPSCs.***

The quantifications of the populations of cells in different cell cycle phases in both dbDNA and oriP-EBNA1 produced iPSCs. An n=3 was analyzed originating from 3 different primary fibroblast cultures with 3 technical repeats per biological replicate. The data was subjected to a Kolmogorov-Smirnov test for normality with the p-value being >0.05 suggesting non-normal distribution of the dataset. As such, a Mann-Whitney U-test was carried out between dbDNA and oriP-EBNA1 iPSCs. * - p-value = <0.05, ** - p-value = <0.01, *** - p-value = <0.001. G0/G1 p-value = 0.068, S phase p-value = 0.317 & G2/M phase p-value = 0.548.

The results demonstrated a positive trend which doesn't reach significance with more dbDNA-iPSCs existing within the *G0/G1* cell cycle phase and likewise that more *oriP*-EBNA1 iPSCs exist within the *G2/M* phase (*Figure 74*).

### 2.1.10. **Summary:**

The dbDNA vector was designed by Touchlight Genetics as a transient expression system. The vector, in theory, should not integrate within its host cells genomic DNA and neither does it have any scaffolding or attachment mechanisms to increase its longevity within such cells. The dbDNA system should, in theory, persist within a cell and express its transgenic cassette whilst being vulnerable to dilution by cell division alongside potential host cell silencing mechanisms.

Early generation expression plasmids by design should share similarities within their functionality to the dbDNA system – with both being transiently designed expression vectors. The two systems do differ structurally, with the dbDNA system cutting out the presence of any bacterial sequences. The dbDNA system contains the desired transgene of interest only, alongside an incredibly short, clinical grade DNA backbone. While the plasmid contains the transgene of interest, alongside a plethora of bacterial DNA backbone.

Our results demonstrated that when transiently transfected into a host cell, that the dbDNA system persists with a greater longevity. This provided the vector with a longer timeframe to express its transgenic cassette which provided us with hope for the dbDNA vectors translational potential into the field of iPSC development. A field where expression plasmids, under a single transfection, do not function to produce iPSCs; instead requiring a minimum of 2 transfections for the induction of pluripotency. Following the reduced functionality of the expression plasmid system, it was modified to contain the *oriP*-EBNA1 DNA sequence. This sequence would provide the plasmid with a scaffolding function, allowing it to tether to host cell chromosomes. This therefore increased the persistence of the plasmid and by virtue, the transgene expression period. This led to the success of the system as it is known today, being utilised in labs globally and in the few clinical trials which have since received approval. The functionality of the plasmid system however was totally dependent on the presence of *oriP*-EBNA1 and its chromosomal tethering capacity. Additional factors have also been added to the standard pluripotency cassette over time, with the aim of improving the efficiency of system within the induction of pluripotency in somatic cells. One common additive was the *shp53* transgene, which is a short hairpin protein for

the *p53* protein. This was determined to improve the efficiency of the reprogramming process but duly also increases the tumorigenic potential of the iPSCs it produces.

All this to say, however, that our data demonstrated that the dbDNA system, which incorporated neither the *oriP*-EBNA1 system, nor the *shp53* transgene could still function effectively within the reprogramming process much to our surprise and excitement. Despite being transient by design, the dbDNA system could function to produce iPSCs reliably, having successfully reprogrammed 11 different primary fibroblast cultures producing stable iPSCs from them all. The dbDNA vector functioned at an equivalent rate to that of the *oriP*-EBNA1 system within this process too, demonstrating a reprogramming efficiency that was not significantly different to that of its well-established counter-part. Moreover, the dbDNA vector could also reprogram and produce stable iPSC colonies from fibroblasts with which it had been previously unachievable with the *oriP*-EBNA1 system.

Taken in its entirety we were able to demonstrate that the dbDNA system functions effectively, reproducibly and was moreover robust in its functional capacity. The *oriP*-EBNA1 vector is utilised globally as aforementioned. For the dbDNA system to therefore match the *oriP*-EBNA1 vector in many functional aspects and even surpass it in certain elements was testament to the novel systems clear potential within this research area and beyond.

We have been able to clearly evidence that the dbDNA system was highly functional within iPSC development. In this chapter we have also demonstrated that the iPSCs produced by this vector system were *bona-fide* by nature. iPSCs should be able to display a dual functionality, incorporating an ability to maintain a pluripotent phenotype, but, when required, also being able to differentiate into cells of all three germ lineages. We were able to demonstrate such properties using several different techniques; ICC and RT-PCR was used to determine the cells pluripotent status and likewise ICC was utilised to demonstrate the cells differentiation capacity as well. All of which were successful in dbDNA produced iPSCs.

We attempted to incorporate the dbDNA system into the reprogramming of cell types outside of fibroblasts. However, the dbDNA system displayed shortcomings in this area, being unable to reprogram neither cells isolated from blood or urine where the *oriP*-EBNA1 system could. The urine reprogramming process was not entirely

reproducible, and the blood reprogramming did yield an '*in-between*' partially reprogrammed phenotype. Duly, these experiments are not the *be-all-and-end-all* the matter and this isn't to say that the dbDNA system won't have successes within these cell types. The data we produced, provided evidence to suggest that the dbDNA system had at least some effect on the adherent PBMCs when reprogramming blood. We did attempt several titrations changing the number of cells plated and the concentrations of DNA nucleofected but with no further success. The process, however, with more work titrating factors such as the transfection process and the number of transfections utilised may still be successful yet for the dbDNA system – this was by no means the end of the road for this vector within this process.

The dbDNA system was also applied to a *xenofree* fibroblast reprogramming protocol. This protocol incorporated the use of animal-free constituents and has been described to be a stumbling point for different vectors being applied to the field. This is an important process as cells that are to be utilised clinically must be reprogrammed using this methodology in order to limit the transfer of animal-borne diseases to the cultured fibroblasts. The aim of our entire project was to utilise this novel dbDNA vector, which can be considered much '*cleaner*' than its successful counterpart, in order to attempt to aid the transition of this cellular technology to the clinic on a more routine basis than exists now. The dbDNA system was successful in this process, producing stable primary colonies that when cultured, underwent positive pluripotent marker staining for *SOX2*. The success demonstrated by ourselves when applying the dbDNA system within this protocol was a great achievement and demonstrated that, pending a number of other mandatory thresholds, that the dbDNA system can be applied within the production of clinical grade iPSCs.

Having demonstrated successes in the development of iPSCs using the dbDNA system, another core aim of the project was to determine differences in the transcription on a global level between the two iPS cell types by using a microarray analysis. Differentially expressed genes (DEGs) were incorporated into several software which then aligned the DEGs with different biological processes. Using this analysis, we were able to demonstrate that DEGs presented by *oriP*-EBNA1 produced iPSCs aligned to largely immune and pro-inflammatory pathways. EBNA1 produced iPSCs demonstrated significant alignments to *Interferon signalling*, *iPSC differentiation* and *DNA damage*. Further work then clarified a distinct trend of an

increased level of *STAT1* transcription secondary to transfection of the *oriP*-EBNA1 vector. Previous literature had demonstrated that an increased *STAT1* transcription and nuclear localisation was initiated following *EBNA1* protein expression. This ultimately increased the cells susceptibility for interferon expression (Wood et al., 2007). In regards to the transcriptional alignment of *oriP*-EBNA1 over-represented genes to cellular differentiation, we demonstrated that *oriP*-EBNA1 produced iPSCs had an increased differentiation phenotype in culture. The dbDNA produced iPSCs demonstrated a more consistent retention of AP stain (which was excluded from areas of spontaneous differentiation) than EBNA1 produced iPSCs. This phenotype was demonstrated to persist irrespective primary fibroblasts cultures used and across several iPSC passages as well. This implied that the differentiation phenotype potentially existed secondary to the vector utilised to develop the iPSCs. This was a key attribute for the dbDNA system as the maintenance of pluripotency and iPSC quality is key to any cells downstream clinical use. Again, secondary to the microarray analysis we demonstrated a clear phenotype of an increased incitement of DNA damage within *oriP*-EBNA1 transfected cells in comparison to dbDNA transfected cells. The result from the comet assay highlighted that *oriP*-EBNA1 transfected cells exhibited an increased level of DNA damage (*Tail DNA, comet length* & *Tail moment*) to a significant degree. Likewise, dbDNA transfected cells retained a greater level of intact DNA (*Head DNA*) to a significant degree. To substantiate this, a second complimentary analysis in relation to DNA damage was undertaken. ICC was carried out utilising a γ-H2AX antibody. The antibody localises and forms foci at areas of variant H2A which develop specifically following a double-strand DNA break and DNA damage. The foci can then be quantified allowing an insight into the proportion of DNA damage being incited within a cell. From this analysis, we demonstrated a significant and consistent increase in the number of foci in *oriP*-EBNA1 transfected cells in comparison to dbDNA transfected cells. Overall, we were able to clearly show that the *oriP*-EBNA1 vector resulted in an increased level of DNA damage incitement/reduced DNA damage repair in comparison to dbDNA transfected cells which maintained a greater level of genomic integrity.

DEGs isolated from the microarray experiment which were over-represented in dbDNA iPSCs were also analysed using several software to determine what biological processes the DEGs interacted within. It was clear from the analysis that the genes

over-represented within dbDNA-produced iPSCs interacted in biological processes mainly associated with the cell cycle. Most of these functionalities were in relation to cell cycle checkpoints and cell cycle progression. We therefore carried out a cell cycle analysis which incorporated the use of PI, a stain which intercalates within cellular DNA to determine different cell cycle phases by using flow cytometry. The results exhibited a clear trend and demonstrated an increased level of cells within the *G0/G1* phase in dbDNA iPSCs and, albeit to a lesser degree, *G2/M* in EBNA1 produced iPSCs. Taken on its own it is difficult to discern too much from the cell cycle experiment without more corroboratory evidence. That said, alterations to the dbDNA vectors transgenic cassette may help to clarify why differences were exhibited in the cell cycle assay. The dbDNA vector didn't incorporate the p53 short-hairpin protein which disrupts and minimises p53 protein expression. The p53 protein will function during cell cycling to inhibit its progression dependant upon the integrity of cellular DNA. If a cell is exhibiting DNA damage, the p53 protein is capable of inhibiting cell cycle progression between G1 and S phases and between G2 and M phases. The incorporation of the short-hairpin protein which inhibited p53 translation in oriP-EBNA1 iPSCs and the lack of such shRNA transgene in the dbDNA system may be the determining factor for such differences exhibited between the two cell types in terms of their cell cycling

### 2.1.11. **Future Work:**

Future work should be in relation to elucidating the dbDNA systems potential to integrate within the genome and in relation to optimising protocols for the systems use in reprogramming primary cell types other than fibroblasts.

There is a great importance behind determining the potential for the dbDNA system to integrate within the genome of the iPSCs it produces. As mentioned, the insertion of vector DNA into the host genome can increase the likelihood for insertional mutagenesis and therefore diminish the potential to utilise the cells and vector for a clinical purpose. In terms of determining the potential for the dbDNA system to integrate within the genome, several experiments could be undertaken. With unlimited time and resources, the only way to unequivocally determine if the dbDNA system has

integrated into iPSC clones is via the use of whole genome sequencing. Whole exome sequencing could be employed to determine the presence of integrations within exons, but this technique ignores the potential for integrations outside of the coding regions of the genome. Hence, whole genome sequencing would be the only test to unequivocally exclude the possibility of integrations within the genome of any iPSC progeny. A more cost-effective method of potentially determining the presence and location of integrations would be to carry out a Linear amplification mediated-PCR or LAM-PCR. The method is often carried out to identify regions of viral DNA integrations via the identification of viral vector flanking genomic sequences. This system often relies on biotin-labelled primers for the viral vectors LTR. The positive DNA sequences are then magnetically captured before several PCR amplifications ensue. The process will have to be adopted to recognise dbDNA-specific sequences but should still function to positively identify any integrations.

For the reprogramming of other primary cell types outside of fibroblasts, this was carried out with little success using the dbDNA system in blood and urine-derived cells. Further optimisation can be undertaken applying different nucleofection programmes, different transfection mechanisms and multiple transfections for example to increase the possibility of deriving iPSC from such primary cells using the dbDNA system.

Moreover, a karyotyping analysis should also be carried out as part of the standard pluripotency assessment to ensure a lack of aneuploidy in the dbDNA produced iPSCs as part of a quality control assessment.

Attempts to clarify any manipulation of STAT1 in an EBNA1-dependent fashion. Work to determine the localisation of *STAT1* following transfection of both vector types to denote any nuclear localisation would help to clarify an increase in activation. Our data suggested no significant difference in transcription of STAT1 was exhibited following vector transfection, however we did not determine the level of phosphorylated STAT1 or its localisation following nucleofection both of which would provide great insight into any differences exhibited between the vectors. Moreover, how the modulation of STAT1 interacts with the differentiation phenotype could also be further elucidated. This could be by *STAT1* over-expression in a pluripotent cell type, such as ESCs for example, and denoting how this affects the cells pluripotent capacity.

In relation to the cell cycle analysis, future work should be to identify the mechanism behind differences between both cell types and their residence in different cell cycle phases. More cells in dbDNA produced iPSCs seemingly resided within the G0/G1 phase of the cell cycle. There was a potential that this was in relation to an increased level of p53 which was present in dbDNA produced iPSCs due to the omission of the *shp53* transgene from the vector. The p53 protein can function to upregulate p21 expression to inhibit the passage of cells through the *G0/G1* phase and into the *S phase*. Therefore, I believe that extra studies into p21 expression will help to provide a clearer image in relation to any differences in cell cycling between the two. Moreover, the microarray also indicated that the dbDNA system shows an increased alignment with the *G2/M progression* biological process. Again, the trend in the cell cycle experiment suggested that a greater proportion of *oriP*-EBNA1 produced iPSCs reside in the *G2/M* phase. It is possible that the *oriP*-EBNA1 produced iPSCs which were more resident in the *G2/M* phase may have been in an arrested phenotype. This in-turn may ultimately result in an increased level of apoptosis. The microarray suggested that the *oriP*-EBNA1 produced iPSC gene expression specifically aligns with the biological process of apoptosis upon analysis using the *GSEA* software. Therefore, further analysis into the level of apoptosis between the two cell types could also be carried out. An *annexin 5* antibody, which will target apoptotic cells specifically followed by a flow cytometry analysis could be undertaken to help substantiate if such reasons can help to explain the mechanism behind the differential cell cycling phenotypes exhibited by both cell types.

## 3.0. Discussion

The process of reprogramming and iPSC development has been established for well over a decade now, following seminal work from the Yamanaka lab in Kyoto, Japan (Takahashi et al., 2007). They were able to describe the principles which ultimately led to the induction of pluripotency within somatic cells and the induction of a new scientific niche which would be studied globally. iPSCs are cell types which function based on two rudimental properties – perpetual self-renewal and a capacity to differentiate into cells of all three germ lineages. Such properties provide iPSCs with a multitude of downstream functional capabilities from organoid development to disease modelling, drug discovery and cell replacement therapy (Rao and Malik, 2012). Despite being discovered over a decade ago, the progression of iPS-cell therapies into a clinical setting has been slow. This has been in relation to factors associated with safety concerns, a lack of efficiency developing a clinical grade iPSC line and issues with the scale-up of production of cells for clinical application.

Genome integrating lentiviruses and onco-retroviruses present significant safety concerns when used therapeutically with insertional mutagenesis being at the forefront of such concerns. Initially, in 2007, the first human iPSC line developed was utilising an integrating retroviral vector by the Yamanaka lab. This method of iPSC development was functionally robust and reliable. In follow up experiments, the Yamanaka lab then developed mice progeny from iPSCs derived from the retroviral system, which harboured known vector integrations within host cell genomic DNA. Around 20% of the mice within the experiment developed tumours secondary to the re-activation of *MYC*-related transgenes (Okita et al., 2007). This was not the first instance in which the use of an integrating retroviral system had adverse secondary effects. In 2003, correction of *X-linked SCID* by ex vivo, retrovirally mediated transfer of the *γc* gene in CD34+ cells was undertaken. Patients treated within this cohort developed leukaemia and/or sarcoma secondary to vector mediated integrations within the CD34+ cells genomic DNA (Hacein-Bey-Abina et al., 2003). Despite the latter of the two providing adverse events not being directly related to iPSCs specifically, such effects secondary to the use of integrative vectors demonstrated the

need for non-genome integrating systems to aid any potential clinical translation of iPSC therapies.

The research into the development of non-integrating reprogramming vectors yielded a multitude of different candidates each with differing levels of success. Research grade iPSC lines were produced using a number of non-integrating systems such as Sendai virus, mRNA, protein transfection and mini-circle vectors (Haridhasapavalan et al., 2019). However, the vector which has received the most widespread use within the field leading to it becoming the *gold standard* of pluripotency induction was the *oriP*-EBNA1 system. The *oriP*-EBNA1 system was first described in the context of iPSC development in 2009, being a system which functioned *footprint free* leaving behind no signs of plasmid or transgene sequence in a host cell with no vector mediated integrations (Yu et al., 2009). The *oriP*-EBNA1 system soon became solidified as the *gold standard* of pluripotency, displaying a functionality that ultimately bettered that of its above-mentioned rivals. The plasmid system was able to reprogram several different primary cell types (*fibroblast, blood, urine*) to iPSCs, conveyed a relatively competitive reprogramming efficiency and required only a single transfection to initiate iPSC development. This makes the plasmid system and its incorporation into the reprogramming niche simple, non-laborious and incredibly cheap. A culmination of all these factors clearly exhibited the reasons why the *oriP*-EBNA1 system was subsequently more readily adopted than any other reprogramming system.

The *oriP*-EBNA1 vector was developed following the failings of previous attempts to apply standard plasmid vectors to reprogramming and iPSC development (Okita et al., 2008; Si-Tayeb et al., 2010).The *oriP*-EBNA1 system was designed to incorporate a standard plasmid DNA backbone with the addition of the *oriP* & EBNA1 sequences. As mentioned, the *oriP*-EBNA1 sequence conferred a tethering function to standard plasmid-based systems. The vector could then be maintained through cell divisions by segregating with the host chromosomal DNA during normal mitotic events. This therefore prevented the dilution of transient plasmid vectors during mitosis within highly proliferating cell systems, such as cells which are undergoing reprogramming (Yu et al., 2009; Hodin et al., 2013; Okita et al., 2013; Drozd et al., 2015). The addition of such sequences increased the plasmids retention, thereby benefiting both the plasmids transgene expression period and its functional capacity for pluripotency induction. The development of this vector system had been a key catalyst in the

progression of a number of clinical trials. Clinical grade iPSC lines produced from the *oriP*-EBNA1 system have, insofar, been the only cells utilised in all clinical trials incorporating iPSC/iPS-based therapies. This clearly demonstrated how vital this system was to the translation of iPSCs to the clinic. Yet, despite this, only 6 clinical trials incorporating iPSCs have been undertaken globally as opposed to 26 using ESCs (Eguizabal et al., 2019). This highlights the nature of the shortfall in terms of iPSC translation; with the reprogramming plasmid vector being implicit within this.

The *oriP*-EBNA1 system is a plasmid-based vector and so incorporates a bacterial DNA backbone and bacterial selection coding sequences. Bacterial DNA has long been recognised as potentiating unwanted, pro-inflammatory effects within host cells, mediated by interactions with *Toll like receptors* (TLRs) (O'Neill et al., 2013). Likewise, the *oriP*-EBNA1 plasmid system also incorporated the EBNA1 protein. Derived from the potent oncovirus EBV, the EBNA1 protein functions synergistically with the *oriP* sequence to physically scaffold the plasmid to host cell chromosomes; a function which has been paramount to the systems success (Yu et al., 2009). However, recent literature suggested that the EBNA1 protein specifically has a number of unwanted interactions within host cells being responsible for potentiating pro-inflammatory responses; evoking the production of reactive oxygen species (ROS) and ultimately inciting DNA damage (Wood et al., 2007; Gruhne et al., 2009). The reprogramming factors incorporated into a vector can also have huge downstream implications with respect to the cell's translational potential. The *oriP*-EBNA1 system incorporated an shRNA against the p53 protein during reprogramming. The p53 protein, often referred to as the *guardian of the genome,* has several functionalities which support its title. Namely, p53 exhibits a rate-limiting function in relation to the cell cycle process. During reprogramming, cells which develop a pluripotent phenotype often become hyperproliferative (Kapinas et al., 2013). During the cell cycle process there are a number of checkpoints at which cells are held to ensure their genomic integrity before being allowed to proceed. The p53 protein is responsible for checkpoints in relation to allowing cells to progress through to the *S phase* and the *M phase* of cell cycling prior to mitosis (Chen, 2016). Cells which are deemed to be unfit to progress, potentially secondary to DNA damage for example, will often undergo apoptosis mediated again by the p53 protein (Zhao and Xu, 2010). It was therefore hypothesised and evidenced that by including the *shp53* sequence within the reprogramming vector, thereby

reducing the level of the p53 protein, that this would improve the reprogramming efficiency (Hong et al., 2009). This was evidenced to be true and independently corroborated (Hong et al., 2009; Marión et al., 2009; Rasmussen et al., 2014). Yet, the greater functional efficiency achieved by incorporating the use of the shRNA for p53 does come at a cost. An increase in tumorigenicity of the iPSCs developed has been reproducibly demonstrated highlighting limitations to the cells downstream potential for clinical use (Chin et al., 2009; Marión et al., 2009; Zhao and Xu, 2010; Merkle et al., 2017). Taken in its entirety, it's clear that the development of the *oriP*-EBNA1 plasmid system has helped to progress iPSC technology and iPS cell therapies from a proof-of-principle technology into a clinically viable treatment modality. However, the number of clinical trials incorporating iPSCs is much lower in comparison to its pluripotent counterpart, ESCs. The *oriP*-EBNA1 vector has been suggested to be at least partly accountable for such a shortfall due to some of the above-mentioned pitfalls of the system being responsible which should not be ignored.

The Doggybone (dbDNA) system is a clinical grade DNA vector first produced and patented by Touchlight Genetics in 2008. The system employs a minimal DNA backbone of just two short 28bp telomeric caps which flank a sequence of interest. The vector was designed as a transient system which does not incorporate any bacterial DNA or the EBNA1 sequence alike. Reprogramming vectors produced using the dbDNA system by Touchlight Genetics utilised the same transgenic cassette as the *oriP*-EBNA1 vector but omitted the inclusion of the shRNA to p53. We therefore had a system with 3 key differences to the plasmid *gold standard*:

1. The dbDNA system negated the inclusion of any bacterial DNA as opposed to the *oriP*-EBNA1 system which contained a bacterial DNA backbone and bacterial antibiotic resistance selection cassettes.
2. The *oriP*-EBNA1 plasmid included the EBNA1 protein which was absent from the dbDNA system.
3. The dbDNA system did not incorporate the shRNA for p53 within its transgenic cassette while the *oriP*-EBNA1 system did.

The over-arching aim of my project was therefore to determine the functional capabilities of the Doggybone (dbDNA) system within somatic cell reprogramming and iPSC production. Following any success in iPSC induction, experiments to elucidate

any differences or benefits conferred by using the bacterial, viral and *shp53*-free dbDNA system should also be undertaken.

### 3.1. Somatic cell reprogramming, pluripotency induction and functional capacity:

The *Doggybone* (dbDNA) expression system is a monomeric, double-stranded, linear, covalently closed DNA construct that can be generated to clinical Good Manufacturing Practise (cGMP) standards. The vector was produced enzymatically beginning with a plasmid template which was ultimately reduced to a desired transgenic sequence inside a covalently closed system – excluding any bacterial DNA (Karbowniczek et al., 2017). The rationale behind using a vector system of this kind within the remit of somatic cell reprogramming was clear when the current vectorology was considered. As detailed earlier, there has been a vast array of reprogramming vectors that have been applied to iPSC development with differing levels of success (Borgohain et al., 2019; Haridhasapavalan et al., 2019). These vector types have largely involved the adoption and manipulation of virus or bacterial plasmid DNA. The current pluripotent *gold standard* vector that has been adopted globally is the *oriP*-EBNA1 plasmid system (Yu et al., 2009).

Standard expression plasmid with a CAG promoter to drive strong transgenic expression was initially adopted into the reprogramming process a year before the *oriP*-EBNA1 system. Despite being successful in iPSC development, its colony forming efficiency was unsustainably low (~0.0001-0.0029%) (Okita et al., 2008; Si-Tayeb et al., 2010). This vector was therefore modified to contain the *oriP*-EBNA1 system, which allowed the plasmid to physically scaffold itself onto host cell chromosomal DNA. The functional advantage of a prolonged expression period has demonstrated a far improved efficiency in comparison to standard expression plasmid and an adoption of this reprogramming method globally (Yu et al., 2009; Okita et al., 2013; Wang et al., 2013). The dbDNA system shares a greater likeness with standard expression plasmid as opposed to the *oriP*-EBNA1 system. The dbDNA vector, in theory, is a non-integrative vector with no additional scaffolding potential. As such, the main functional concern when applying dbDNA to iPSC reprogramming in this project

was surrounding the transiency of the system and a potential for a reduced functionality (Okita et al., 2008; Si-Tayeb et al., 2010; Haridhasapavalan et al., 2019).

The dbDNA system however was successful when applied to the McKay lab protocol for reprogramming and iPSC development – reprogramming a total of 11 different primary fibroblast cultures (Hawkins et al., 2016). Our data demonstrated that the dbDNA vector was robust and reproducible in the induction of pluripotency within fibroblasts. The system functioned with a higher success rate than the *oriP*-EBNA1 system and with a non-significantly different reprogramming efficiency. The literature suggested that a reprogramming efficiency of <0.01% was a very low reprogramming efficiency, with 0.01-0.1% considered as low and >0.1% considered as a high reprogramming efficiency (Haridhasapavalan et al., 2019). The *oriP*-EBNA1 system was consistently denoted to express a reprogramming efficiency that was considerably lower than its fellow reprogramming counterparts (Yu et al., 2009; Rao and Malik, 2012; Schlaeger et al., 2015; Borgohain et al., 2019). In a direct comparison, reprogramming experiments were carried out within the Daley lab incorporating the *oriP*-EBNA1, Sendi virus and mRNA reprogramming methods. The *oriP*-EBNA1 system displayed the lowest average reprogramming efficiency (0.013%) before Sendi at 0.77%, with mRNA having the highest reprogramming efficiency at 2.5% (Schlaeger et al., 2015). When projected into the efficiency categories outlined by Haridhasapavalan, the *oriP*-EBNA1 system has a *low* reprogramming efficiency and comparatively to other non-integrating vectors it was also considerably lower. This could be due to several reasons. Yet, the mRNA system, displayed what seemed to be a high reprogramming efficiency, but the protocol associated with this method of reprogramming was time-consuming, arduous and difficult to replicate. The mRNA system relies on daily transfections of its reprogramming factors and the use of chemicals to supress the transfected cells immune response to the system. The efficiency of iPSC production using the *oriP*-EBNA1 system was objectively lower to that of mRNA, however, the reprogramming process was much more simplistic and less time-consuming, owing only to a single transfection. A delicate balance clearly existed in relation to reprogramming efficiency and the workload and reproducibility of the protocol. However, despite having a reprogramming efficiency lower than other non-integrating vectors, the *oriP*-EBNA1 system has made the biggest steps, being adopted globally and utilised in all approved iPSC clinical trials (Eguizabal et al., 2019).

The reprogramming efficiencies achieved by both the dbDNA and *oriP*-EBNA1 system in our experiments was *0.13%* and *0.16%* respectively. According to the Haridhasapavalan classifications the dbDNA system would be deemed to have a *high* reprogramming efficiency (Haridhasapavalan et al., 2019). However, in relation to other non-integrative reprogramming systems, its reprogramming efficiency was on the lower end of the scale. That said, it was clear that the efficiency at which a vector can reprogram was not the cornerstone of the technologies potential to be translated to the clinic – as the vector with one of the lowest relative efficiencies was recognised as the *gold standard* of the field. As previously stated from our dataset, the dbDNA system displayed a success rate that was greater (*92%*) than the *oriP*-EBNA1 vector (*84%*) in relation to the 12 different primary fibroblast cultures which underwent reprogramming using both vectors. In context with other reprogramming vectors the *oriP*-EBNA1 system displayed a high reprograming success rate and reproducibility within the reprogramming of fibroblasts. The *oriP*-EBNA1 plasmid displayed a *93%* success rate as opposed to a *27%* and *94%* success rate for mRNA and Sendi respectively in experiments outlined by the Daley lab (Schlaeger et al., 2015). In relation to the literature and by the experiments carried out by Schlaeger, the dbDNA system displays a competitively high success rate when compared to both integrating and non-integrating vectors. Despite differences between the vectors structurally, the dbDNA and *oriP*-EBNA1 systems were incorporated into identical protocols to yield the results obtained during this project. A novel facet of the literature published by the Daley lab was that they carried out a survey across 149 laboratories who were involved in iPSC research, 94 of which the author determined to be '*experienced*' within the field. The survey was used established the success rate for the *oriP*-EBNA1, mRNA and Sendi vectors in the reprogramming of fibroblasts. Across 22 labs, the mRNA method had a *59%* success rate, Sendi across 35 labs had a *97%* success rate and the *oriP*-EBNA1 system across 21 different, '*mixed ability'* laboratories displayed a *100%* success rate (Schlaeger et al., 2015). This demonstrated that, irrespective of experience, the reconstitution of pluripotency using the *oriP*-EBNA1 system was a reproducible process. Inferences from this were that the protocol in relation to iPSC development using the *oriP*-EBNA1 system was basic and simplistic enough to be carried out in labs globally without error. As mentioned, the dbDNA system incorporated the same protocol as the *oriP*-EBNA1 system. Thereby, in theory, if

incorporated into iPSC labs globally the dbDNA system could also work flawlessly irrespective of lab user experience. The literature substantiated the idea that the dbDNA system not only functionally matched the *oriP*-EBNA1 system but could also be potentially implemented in labs globally without fault.

Moreover, one of the omitted factors from the dbDNA system was in relation to its lack of *shRNA* for the p53 protein which was present in *oriP*-EBNA1 iPSCs. The shRNA will interfere with p53 translation and protein expression following *shp53* expression. This results in a reduction/loss of p53 function during the reprogramming process and iPSC development. The literature surrounding the function of p53 during reprogramming and its resulting effects on its iPSC progeny is conflicting. It was first theorized and evidenced in the Yamanaka lab in 2009 that the interactions between p53 and p21 resulted in a reduction in reprogramming efficiency. Moreover, the publication suggested that using the knock-down of p53 during reprogramming experiments could still yield integration-free iPSCs but at a greater reprogramming efficiency (Hong et al., 2009). Literature has since been published supporting the theory that p53 expression demonstrated a suppressive function in terms of reprogramming success and efficiency (Marión et al., 2009; Yi et al., 2012; Rasmussen et al., 2014). Experiments outlined in the Claussen lab described that reprogramming with the p53 *shRNA* yields around a 6-fold increase in colony production bringing the reprogramming efficiency, using *oriP*-EBNA1 plasmid, to ~0.12% (Rasmussen et al., 2014). The results mirror reprogramming efficiencies obtained when using the plasmid system with a *shp53* transgene as outlined in our results - with reprogramming efficiencies persisting at around 0.16%. The work published by Claussen also detailed that they see no compromise to the integrity of cellular DNA by using flow cytometry in conjunction with a γ-H2AX antibody at regular timepoints in fibroblasts following the initiation of reprogramming. However, conflicting reports clearly detailed that the loss of p53, despite increasing the reprogramming efficiency, compromised the genomic integrity of the iPSC progeny produced when reprogramming in this manner. The Blasco lab provided a detailed report, evidencing that during the reprogramming of fibroblasts to iPSCs, that the *shRNA* for p53 abrogated apoptosis onset, as demonstrated by a flow cytometry analysis using an *Annexin A5* antibody. Moreover, the publication also outlined that the cells during and following successful reprogramming with a shRNA for p53 demonstrated an increased

level of DNA damage. ICC using a γ-H2AX antibody and foci quantification, accompanied by chromosomal end-to-end fusion analysis was undertaken. Both of these types of analysis detected increased levels of DNA/chromosomal damage in cells incorporating *shRNA* in comparison to control cells (Marión et al., 2009). Despite some reports suggesting that reprogramming with p53 results in no genetic aberrations, it was seemingly more broadly accepted that the process can result in the propagation of iPSCs with DNA damage and can limit the cells clinical applicability. The premise of our project was in relation to the dbDNA systems ability to produce safer iPSCs. It was determined that the *shRNA* for p53 should therefore be omitted from the dbDNA sequence. What was surprising about the results obtained when reprogramming using the dbDNA sequence in comparison to the *oriP*-EBNA1 system was that the Doggybone vector was able to reprogram and produce iPSCs at an efficiency that was not significantly different to that of the EBNA1 vector. This result requires further work to truly elucidate the mechanistic reasoning behind it. However, what was clear was that the dbDNA system was able to reprogram and produce iPSCs at a greater basal rate than the plasmid-based alternative as studies detailed a large-scale increase in reprogramming efficiency following the inclusion of the p53 *shRNA* to levels of that exhibited by the dbDNA vector without the shp53 sequence (Rasmussen et al., 2014). A hypothesis for this could be that, as demonstrated by ourselves, the dbDNA vector incited less DNA damage following cellular transfection than the *oriP*-EBNA1 vector; with EBNA1 being well established in the induction of DNA damage (Gruhne et al., 2009). It is therefore within reason to hypothesise the grounds behind the increased basal reprogramming efficiency exhibited when using the dbDNA system was in relation to evidence that the system initiates less DNA damage. In turn the dbDNA system would potentiate a far reduced p53 response and subsequent limit on cellular reprogramming than what would be exhibited when utilising the *oriP*-EBNA1 plasmid without the *shp53*. As mentioned, however, large-scale experiments would be required to elucidate the real mechanistic reasoning behind this result.

Taken in its entirety, the results obtained by ourselves in conjunction with the literature demonstrated that the ability of dbDNA system to function in iPSC development was an extraordinary result. The vector, according to ourselves and the literature, functioned with a relative equivalence to that of the *oriP*-EBNA1 system. On a broader

scale, the reprogramming efficiency may not be high, relative to other reprogramming systems, but neither was that of the *oriP*-EBNA1 system which has still managed to be recognised as the pluripotent *gold standard* globally. However, the dbDNA system structurally contained no *shp53* transgene and was designed to be a non-integrative, transient system with no scaffolding ability. Therefore, despite these factors, which previous literature has described as having fundamental effects on the efficiency of iPSC colony formation, the dbDNA system was still able to function at a similar rate to that of the *oriP*-EBNA1 system. This was an incredibly exciting achievement which places the dbDNA system in good stead for further development within the reprogramming field.

### 3.2. Increased retention of the dbDNA system and interactions with host cell innate immunity:

So far, I have demonstrated that the dbDNA system can reproducibly reprogram fibroblasts producing *bona fide* iPSCs. The vector did so without the need of any scaffold attachment proteins to improve its retention, as the *oriP*-EBNA1 system did. Previous literature provided substantiating evidence that a standard expression plasmid was not functionally apt within reprogramming, displaying a low success rate and efficiency (0.0001-0.0029%) (Okita et al., 2008; Si-Tayeb et al., 2010; Haridhasapavalan et al., 2019). To determine and substantiate the novelty of the dbDNA vectors functionality within iPSC production, a large-scale re-programming experiment was carried out. This experiment employed the dbDNA vector, *oriP*-EBNA1, an expression plasmid (without *oriP*-EBNA1) and a GFP plasmid for a negative control. The results from that experiment demonstrated viable, AP positive colonies were produced from dbDNA and *oriP*-EBNA1 nucleofected fibroblasts and no AP positive colonies in the expression plasmid or negative control. The results from this experiment provided further evidence supporting the novel nature of the ability of the dbDNA system to induce pluripotency in fibroblasts. The *oriP*-EBNA1 null plasmids employed in this experiment were polycistronic vectors employing the same factors as both the dbDNA and *oriP*-EBNA1 systems, with each transcript being separated by a *2A* sequence. Previous literature has suggested a need to employ multiple

transfections for the successful function of expression plasmid systems in iPSC development, owing to the vectors transient nature (Okita et al., 2008). This was true also of polycistronic plasmids which were demonstrated to produce iPSCs from human dermal fibroblasts again at the expense of more than one transfection timepoint (Si-Tayeb et al., 2010). Therefore, the lack of AP-positive iPSC development from expression plasmids within the McKay lab protocol, which incorporated a single transfection was understandable. During reprogramming it had been demonstrated that strong transgenic expression was required for a minimum of 12 days in order to induce the development of primary iPSC colonies (Brambrink et al., 2008). It was clear from our results in conjunction with previous literature, that expression plasmid was too transient in nature to fulfil a sustained transgenic expression over this required time period following a single transfection; multiple transfections were necessary. However, the dbDNA system was clearly capable of expressing its transgenic cassette beyond the stipulated 12-day period and could successfully induce iPSC development from a single transfection, despite being designed as a transient system.

The ability of the dbDNA vector to produce iPSCs, concurrent to the expression plasmids failure to do so, highlighted the novelty of Doggybones functionality within the context of reprogramming. Such data was indicative of a difference within host cell recognition of the dbDNA system in comparison to traditional plasmid vectors. The dbDNA vector seemingly displayed an increased retention in comparison to the plasmid-based system in order to be able to reconstitute pluripotency within somatic cells. Results procured by ourselves indicated that the dbDNA system had a more prolonged expression timeframe and an increased intensity of expression over its plasmid counterpart. Both the dbDNA system and expression plasmid are transient vectors by design, therefore the increased longevity of expression within the dbDNA system suggested a greater retention of this vector type within cells, or inversely, a more prompt dilution/silencing of the expression plasmid. Both vector types shared an identical transgenic sequence and subsequently produce the same protein (GFP) too. The difference between the two vectors was structural, being in relation to the presence/absence of a bacterial DNA backbone. The expression plasmid was the template structure used when producing dbDNA from *rolling circle amplification* (RCA) and so was structurally very similar to the Doggybone system. The only difference between the two systems was that the bacterial DNA backbone outside of the

transgenic sequence in the plasmid was omitted from the Doggybone vector (Karbowniczek et al., 2017). It was therefore reasonable to rationalise that the difference in retention between the two systems was dependent upon the remaining bacterial DNA in the expression plasmid and its interaction within the transfected host cell; an interaction that the bacterial DNA-free dbDNA system was able to bypass.

The enzymatic method of producing the novel dbDNA vector relies on the presence of a plasmid starting template but ultimately produces a product that is devoid of plasmid backbone DNA, CpG motifs and antibiotic-resistance genes that are required for the bacterial propagation of plasmid DNA (Karbowniczek et al., 2017; Allen et al., 2018). First hypothesized in 1989, much evidence has been developed to underpin the presence of an inherited cellular recognition domain (Pattern Recognition Receptor – PRR) capable of detecting pathogenic associated molecular patterns (PAMPs) (Dempsey and Bowie, 2015). Moreover, the discovery of Toll like receptors (TLRs) provided a further mechanistic foundation for pathogenic detection and ultimately an inflammatory cellular response (O'Neill et al., 2013). In 2000, a PRR was implicated in the detection of double-stranded (ds) DNA, TLR9. This toll like receptor functions in response to the abundance of unmethylated DNA (CpG motifs) present in bacterial DNA before inducing cytokine expression whose function ultimately terminates with an increased propensity for interferon expression (Hemmi et al., 2000). The incitement of such an inflammatory response was likewise associated with an increased level of plasmid removal and transgene silencing - whereby the plasmid was still present but its expression becomes redundant (Qin et al., 1997; Yew et al., 2000). Work has been carried out to determine the ability of the TLR9 system to recognise and potentiate a downstream inflammatory response in relation to the dbDNA vector (Allen et al., 2018). Experiments detailed by the Savelyeva lab incorporated the use of a *HEK293T* cell line over-expressing the TLR9 bacterial DNA recognition receptor. When the TLR9 over-expression line interacted with a piece of DNA yielding a positive result the cell line produced a secreted alkaline phosphatase (SEAP) which could be quantified. The results demonstrated clear experimental evidence that the dbDNA vector bypassed recognition by TLR9 receptors when transfected into the overexpression *HEK293T* cells. While plasmid DNA potentiated close to 5-fold increase in SEAP levels – indicating activation and processing of the plasmid by the TLR9 receptor. Moreover, this effect was quashed following the inhibition of TLR9 function suggesting the

plasmid processing was TLR9 specific (Allen et al., 2018). It was reasonable to therefore rationalise that the reduced retention and intensity of plasmid expression within our experiments was due to the processing of the plasmid bacterial DNA backbone via TLR9 systems. This, in turn, would result in an increased potential for transgene silencing in combination with plasmid dilution which ultimately reflects a reduced functionality. From this initial dataset obtained by ourselves in conjunction with the literature, it was clear that the dbDNA system potentiated a much-reduced pro-inflammatory cellular environment in comparison to expression plasmid. This conveys positive functional benefits to the Doggybone vector in terms of its longevity of expression and increased functionality within the reprogramming niche.

The literature has demonstrated that the *oriP*-EBNA1 plasmid system is an incredibly robust vector in terms of its ability to reprogram fibroblasts alongside a number of cell types from different somatic sources (Zhou et al., 2012; Okita et al., 2013). The increased efficacy and efficiency of the vector was, without doubt, due to the nature of the scaffolding abilities incited by the inclusion of the *oriP*-EBNA1 system (Yu et al., 2009). The literature detailed how plasmid systems without any tethering abilities were more susceptible to a reduced functionality secondary to increased vector dilution and/or transgene silencing. It was ubiquitously recognised within the literature that plasmid dsDNA was processed by a host somatic cell via the TLR9 system; initiating a proinflammatory response (Yew et al., 2000; Bauer et al., 2001; Allen et al., 2018). It is therefore reasonable to postulate that the pro-inflammatory response to the plasmid was the mediator of its reduced functionality within somatic cells. However, now with the inclusion of the *oriP*-EBNA1 system within the same plasmid framework, this provided the vector with an increased retention and as such an increased efficiency in terms of its scope for iPSC development. Yet, it is important not to overlook the fact that such vectors will still be potentiating the same inflammatory response as before, however, with the new scaffolding system this reduced the potential for plasmid dilution. More, the host cell will be subjected to a persistent, unwavering pro-inflammatory signature.

The above data and supporting literature provided the foundations for the hypothesis that upon developing iPSCs from both the dbDNA and *oriP*-EBNA1 systems, that one difference between these two cell types could be in relation to the incitement of a pro-inflammatory microenvironment on a vector specific basis. To begin to clarify any

differences in relation to inflammation and beyond, it was determined that a large-scale experiment was the optimal approach and as such a microarray analysis was undertaken. This would provide information on the cell's probed transcriptomics on a more global level as opposed to trying to pinpoint specific genes in a more bias manner.

The microarray analysis yielded global transcriptional changes between *oriP*-EBNA1 and dbDNA iPSCs. This was reproducibly demonstrated using R studio to produce a correlogram, principle component analysis and volcano plot which demonstrated such changes. Further analysis incorporating over-represented probes for dbDNA and *oriP*-EBNA1 iPSCs alike into a heatmap was undertaken. This was carried out alongside the pluripotent gold standard ESC and it was clear that iPSC over-represented probes were being manipulated in a vector specific manner. The data demonstrated that the *oriP*-EBNA1 over-represented probes were significantly downregulated in dbDNA iPSCs. The expression of the same probes in ESCs demonstrated a more similar expression pattern to that of dbDNA iPSCs than that of *oriP*-EBNA1 iPSCs; determined hierarchical clustering. This suggested that probes that were over-represented in *oriP*-EBNA1 iPSCs were largely done so specifically within that cell type and not in dbDNA iPSCs or ESCs. It was therefore reasonable to hypothesize that such probes were being over-expressed secondary to the internal processing of reprogramming vector adopted. The same analysis for dbDNA iPSC over-represented probes provided a similar outcome, with several probes being manipulated in a vector dependent manner.

To have iPSCs produced from the same primary fibroblast cultures using different vectors providing different global transcriptomic profiles was not uncommon. Work published from the Wu lab analysed the global transcriptomics of iPSCs produced using 6 different reprogramming vectors (*episomal plasmid, mRNA, mRNA + miRNA, minicircle vector, lentivirus & Sendi virus*) and demonstrated that key global differences existed within each vector type in relation to each other and in relation to ESCs using a PCA (Churko et al., 2017). Moreover, the PCA also denoted a relatively vast level of variation amongst the global genomes within iPSCs produced using the *oriP*-EBNA1 system alone highlighting the vectors inability to accurately recapitulate pluripotency within fibroblasts. The Wu lab also quantified such differences using a correlogram and demonstrated key transcriptional differences between iPSCs

produced using different vectors. It was therefore evident that the vector of choice can influence variability amongst iPSCs produced from the same fibroblasts (Churko et al., 2017). For ourselves what was more illuminating was how the differentially expressed genes interacted within different biological processes. This type of analysis provided detail and insight as to what specific biological processes our differentially expressed probes were interacting with and what their potential effects on the host cells intracellular environment were.

The projection of probes over-represented within different biological processes using the Reactome database suggested a clear alignment of *oriP*-EBNA1 probes within '*immune system'* processes; a common theme that continues to be portrayed within the unbiased analysis across 3 different gene ontology (GO) web pages. Moreover, within such '*immune system'* processes, an increased propensity for '*cytokine signalling in the immune system*' and '*interferon-related signalling pathways'* were again a commonality highlighted in the online facilities. It was apparent from the literature that bacterial DNA present in plasmid vectors was processed through the TLR9 receptor, which the dbDNA system managed to bypass (Allen et al., 2018). Following TLR9 receptor stimulation by CpG-DNA molecules, TLR9-endosomes traffic to lysosome-related organelles before a signalling cascade ultimately triggers the induction of type 1 interferon expression (Kawasaki and Kawai, 2014). What was clear from our dataset produced from different GO online software, was that a phenotype associated with the mechanistic pathway downstream from TLR9 activation was displayed. '*Endosomal* and *phagosomal pathways'* alongside *'interferon signalling pathways'* were commonalities within our unbiased analysis and demonstrated the potentiation of an inflammatory intracellular signature. It has been reproducibly evidenced within scientific literature that interferon expression also manipulates the upregulation of Major Histocompatability Complex (MHC) related genes (Gaczynska et al., 1993; Steimle et al., 1994; Thelemann et al., 2014). This was something also detected within our analysis with MHC gene over-expression commonly being detected. The initial results from the microarray provided clear evidence that a pro-inflammatory signature was transcriptionally detectable in *oriP*-EBNA1 produced iPSCs in comparison to dbDNA-produced iPSCs mediated by the presence of bacterial DNA in the plasmid system.

Residual bacterial DNA present in plasmid vectors such as the *oriP*-EBNA1 system has been demonstrated within the literature to potentiate inflammatory responses which mimic those seen within our dataset. The Park lab carried out detailed and stringently controlled RNA-sequencing experiment on fibroblasts which had been subjected to somatic cell reprogramming using *oriP*-EBNA1 vectors (Tanaka et al., 2015). The idea of their experiment was to carry out RNA-Sequencing on fibroblast samples 3 days post-transfection of OSKM *oriP*-EBNA1 vectors. The experiment was carried out in parallel with OSKM-negative plasmid vectors (consisting of plasmid backbone only) and a negative control of un-transfected parental fibroblasts. The results following analysis using *Gene Ontology* software demonstrated that in comparison to parental fibroblasts, the most over-represented genes in both the OSKM positive and negative plasmid transfected fibroblasts was in relation to "*Type 1 interferon-mediated signalling pathway*", "*Cellular response to type 1 interferon"* and *"response to type 1 interferon"* which were all aligned with a high degree of specificity. The research article suggested that the triggering of type 1 interferon pathways by both the OSKM plasmid system and the empty plasmid system was a general cellular response to foreign DNA and not to the OSKM transgenes per se, as substantiated by the identical response in the OSKM null plasmid system too (Tanaka et al., 2015). Further analysis undertaken by ourselves into interferon signalling highlighted by the microarray analysis yielded a heatmap of interferon-related genes almost all of which were transcriptionally over-represented in *oriP*-EBNA1 produced iPSCs only. Our dataset provided clear evidence that interferon related gene expression was being manipulated and transcriptionally over-represented in *oriP*-EBNA1 produced iPSCs alone. This suggested that the *oriP*-EBNA1 system was potentiating intracellular responses secondary to its structure with which the dbDNA system did not. This had been suggested to be purely in relation to the presence of pro-inflammatory bacterial DNA present in the plasmid system insofar. However, the bacterial DNA element may not be solely responsible for the potentiation of interferon related genes. Experiments carried out in the Young lab demonstrated that the EBNA1 protein itself was likewise able to elicit an intracellular pro-inflammatory response (Wood et al., 2007). The lab was able to generate a carcinoma cell line which over-expressed the EBNA1 protein. Such cells were found to result in the activation of *STAT1* hyper-transcription which resultingly promoted a greater level of *STAT1* protein nuclear localisation and

activation using ICC. This resulted with the cells demonstrating a greater sensitisation to interferon expression and also the induction of *MHC* gene expression as demonstrated via western blotting (Wood et al., 2007). Taken as a whole, the literature suggested a synergistic potentiation of gene sets in relation to inflammation and interferon expression with both the plasmid DNA and EBNA1 protein expression being complicit in the unwanted secondary responses. Responses which were bypassed by the dbDNA system, much to its benefit.

Within our dataset, another commonality in the GO comparisons of over-represented probes in *oriP*-EBNA1 iPSCs was '*Cellular differentiation*' and '*Epithelial-to-Mesenchymal Transition*' which represented a reduced ability for iPSCs to maintain their pluripotent capacity. This phenotype was noticed prior to the RNA-sequencing experiment where, as mentioned, certain primary fibroblasts reprogrammed with the dbDNA system produced stable iPSC lines where the *oriP*-EBNA1 system failed (producing lines which ultimately spontaneously differentiated). Moreover, whilst culturing the cells on a routine basis, it became clear that *oriP*-EBNA1 iPS cells did persistently display an increased level of spontaneous differentiation in comparison to dbDNA generated iPSCs. To quantify this, iPSCs produced when using both the dbDNA and *oriP*-EBNA1 vectors in concomitant experiments were simultaneously analysed using an AP stain which specifically interacted with iPSCs only, excluding any areas of differentiation. The result evidenced that across 3 different iPSC lines and numerous passages that a differentiation phenotype persisted in *oriP*-EBNA1 derived iPSCs and not in dbDNA produced iPSCs.

Displaying a differentiation potential is a critical function of iPSCs. Maintenance of a self-renewal phenotype is important but the therapeutic benefit of this cell type is derived from its ability to differentiate into any cell of the three germ lineages (Takahashi and Yamanaka, 2016). However, it has been established that cells produced by the dbDNA system were not '*locked*' into pluripotency, as we have demonstrated that iPSCs produced by this vector were capable of differentiating and forming cells of the mesoderm, endoderm and ectodermal lineages. The lack of differentiation displayed during normal culturing conditions was not a symptom of a lacking capacity for any differentiation, more it was seemingly an incapability of the *oriP*-EBNA1 system within the iPSCs it produces. Reaffirmation of the increased capacity for spontaneous differentiation was demonstrated following a heatmap

analysis of early differentiation markers from our microarray analysis. Such probes were over-represented in *oriP*-EBNA1 iPSCs in comparison to dbDNA iPSCs and ESCs. Within the literature, it was more recently established that following an upregulation of interferon expression, secondary to the recognition of foreign DNA through the above-mentioned mechanisms, that iPSCs could not maintain pluripotency and ultimately undergo unwanted spontaneous differentiation. The Oever research lab demonstrated that, in fact, the initiation of any aspect of the anti-viral response was incompatible with pluripotency. The over-expression of interferon-related genes resulted in the diminished transcriptional expression of *KLF4*, *SOX2* and *OCT4*. Such iPSCs were then incapable of maintaining pluripotency and were instead pushed to spontaneously differentiate towards a mesodermal lineage (Eggenberger et al., 2019).

The mechanism substantiated within the literature helped to coalesce the transcriptional phenotype demonstrated within our iPSC culture with the microarray results we likewise developed. Our results indicated that the potential increased level of differentiation observed within *oriP*-EBNA1 iPSCs was secondary to the processing of its volatile foreign DNA. Within the discussion we have demonstrated clearly that the bacterial DNA backbone along with EBNA1 protein expression have key roles in the potentiation of a pro-inflammatory cellular environment and therefore the manipulation of *interferon-related genes*. Our dataset clearly laid out evidence to suggest that the pro-inflammatory phenotype was maintained and persisted within iPSCs and has clear, unwanted secondary effects in relation to the increased potentiation of a differentiation phenotype. The dbDNA system, free of bacterial DNA and the EBNA1 sequence bypasses such secondary unwanted effects in terms of inflammation and the primed differentiation phenotype which ensues.

### 3.3. EBNA1 protein and increased susceptibility for DNA damage:

Following the microarray, GO terms associated with '*DNA damage responses'* and responses to '*oxygen containing compounds'* were highlighted within our analyses in *oriP*-EBNA1 iPSCs. This was the rationale behind examining the potential for both the dbDNA and *oriP*-EBNA1 system to induce DNA damage within a transfected cell type.

A comet assay in *HEK293T* cells was determined to be appropriate for such analysis. As such, *oriP*-EBNA1-eGFP vectors and dbDNA-eGFP vectors were separately transfected into *HEK293T* cells. GFP expression was qualified after 24 hours before the cells were analysed amongst untreated (negative control) cells and cells treated with 200µM hydrogen peroxide (positive control) (Uryga et al., 2015). Images were collected and analysed using CaspLab software. The results demonstrated an increased level of DNA damage within cells transfected with the *oriP*-EBNA1 vector. The amount of *Head DNA* was lower in comparison to the dbDNA system, alongside an increased level of *tail DNA* which suggested a significantly increased level of damage. Moreover, the *comet length* was higher in *oriP*-EBNA1 cells and also the *tail moment*. Our results provided experimental evidence to suggest that the *oriP*-EBNA1 vector was inducing DNA damage in a way that the dbDNA system seemingly avoided. To consolidate this dataset, ICC was carried out incorporating a γ-H2AX antibody. The antibody functioned based on the detection of the phosphorylated form of variant histone H2AX (γ-H2AX), which occurred specifically at sites of DNA double-strand breaks (DSBs) (Redon et al., 2009). The same experimental procedure was carried out on fibroblasts transfected with dbDNA/*oriP*-EBNA1-eGFP. However, as the GFP protein was retained by transfected cells following ICC processing, this permitted the specific targeting and quantification of γ-H2AX positive foci in transfected cells only. This would therefore permit a more precise account of the level of DNA damage being inflicted and/or the lack of DNA repair initiated within each vector transfected cell-types specifically. The results from this experiment demonstrated that *oriP*-EBNA1 plasmid transfected cells demonstrated a consistently greater level of positive γ-H2AX foci in comparison to the dbDNA system. Taken in their entirety, the results of the comet assay and γ-H2AX analysis demonstrated that the *oriP*-EBNA1 vector could incite a greater level of DNA damage and/or inhibited the cellular repair mechanisms to a significantly greater degree in comparison to the dbDNA system.

As previously stated, the dbDNA system shared the same transgenic cassette as the *oriP*-EBNA1 plasmid system, but was devoid of the plasmid backbone and *oriP*-EBNA1 sequences (Karbowniczek et al., 2017). The mechanism behind the induction of an increased level of DNA damage within *oriP*-EBNA1 transfected cells and not within dbDNA transfected cells must be apropos to something within the remaining bacterial DNA/*oriP*-EBNA1 system. Previous literature published from the Masucci lab

presented evidence that the EBNA1 protein was implicated within the induction of an increased production of reactive oxygen species (ROS) which ultimately manifested as an increased level of double-stranded DNA breaks (DSB) (Gruhne et al., 2009). The lab produced a tetracycline inducible, constitutively expressing EBNA1-protein cell line and so could manipulate its expression within the same cell. Using DCFDA, a ROS specific stain, in conjunction with flow cytometry, Gruhne demonstrated that EBNA1-expressing cells exhibited a higher level of ROS production. Then, the lab group utilised similar experimental procedures employed by ourselves. A comet assay was undertaken which demonstrated that cells with a greater level of EBNA1 protein production exhibited a much-increased comet length in comparison to control cells; indicating a greater level of DNA damage. This was furthermore substantiated by ICC analysis incorporating a γ-H2AX antibody. The publication demonstrated positive staining within EBNA1 positive cells, but the foci were not however quantified. The reversal of such effects in EBNA1 positive cells were demonstrated to be reversed following the administration of ROS-scavengers such as the glutathione peroxidase mimetic *ebselen* and citric acid. The mechanistic analysis put forward by the paper's author was in relation to the activation of NADPH oxides via the transcriptional activation of the *NOX2* gene which was specifically induced by EBNA1 protein expression. The activation of NADPH oxidases ultimately led to the induction of DNA damage in cells which occurred in an EBNA1-dependant manner (Gruhne et al., 2009). Our results provided key indications that the *oriP*-EBNA1 system was inciting DNA damage/interfering with cellular repair mechanisms in a vector-specific manner. The increased prevalence of genomic instability within iPSCs induced by the reprogramming vector could ultimately prohibit the clinical translation of such cells and scupper any therapeutic use (Marión et al., 2009; Sullivan et al., 2018). The initiation of DNA damage by the *oriP*-EBNA1 vector, alongside the inhibition of p53 expression can increase the prevalence of iPSCs harbouring unwanted genetic mutations and modifications. As prior mentioned, this could have damaging effects in terms of an increased oncogenic potential of such iPSCs (Gore et al., 2011; Zhang et al., 2013). This presented the development of much more stable iPSCs using the dbDNA system, free from the volatile EBNA1 sequence and also the shRNA for p53. Many of the proposed issues demonstrated by ourselves and within existing literature seemingly occur secondary to the *oriP*-EBNA1 vector structure and the interactions it has within

the host cell. Yet, despite this, the *oriP*-EBNA1 system has become the gold standard vector for pluripotency induction. Here, we demonstrated that the dbDNA system can match the EBNA1 vector in terms of its functionality and can moreover bypass several unwanted secondary responses, such as increased DNA damage, spontaneous differentiation and inflammation. The dbDNA vector could therefore become the first of a newer, safer generation of reprogramming vectors and could potentially help to transition this therapy into the clinic more readily than what was being achieved with the *oriP*-EBNA1 system.

### 3.4. dbDNA over-expressed genes and cell cycle analysis:

Following the analysis of gene probes that were over-represented in dbDNA iPSCs in comparison to *oriP*-EBNA1 iPSCs it was clear that GO terms in relation to cell cycling were common results. From this we decided to carry out a cell cycle analysis between dbDNA and *oriP*-EBNA1 produced iPSCs using a propidium iodide (PI) nuclear stain. The results demonstrated a trend existing with dbDNA-produced iPSCs seemingly having a greater proportion of cells existing within the *G0/G1* phase in comparison to the *oriP*-EBNA1 system. Whilst the *oriP*-EBNA1 produced iPSCs displayed a greater proportion of cells within the *G2/M* phase. Typically, a highly proliferative pluripotent cell type such as iPSCs would have a more abbreviated cell cycling with a reduced *G0/G1* but normal cycling periods for the remaining phases (Kapinas et al., 2013). The dbDNA system, as mentioned, did not employ the use of the *shRNA* to p53, as the *oriP*-EBNA1 system did. The p53 protein, commonly referred to as the '*guardian of the genome*' has functionalities within cell cycling being responsible for initiating cell cycle arrest and DNA repair (Chen, 2016). Upon detecting genetic abnormalities, the p53 protein, in synergy with p21, was able to halt the cell cycle to prevent such damaged cells from progressing through either the G0/G1 or S phase and ultimately propagating such damage. Ultimately, if a cell type has incorporated the shp53 system, the cell cycling and division of such cells can continue unrestricted and in spite of the presence of potential DNA damage. Ultimately, with the dbDNA system permitting the expression of the p53 protein, it was reasonable to hypothesise that major differences

within cell cycling changes which exist between the two cell types may be a consequence of the presence of this factor.

### 3.6. Reprogramming of different primary cell types:

The dbDNA system was successful in its ability to reprogram fibroblasts using both a proof-of-principle and a *xenofree* protocol. This was an extraordinary achievement for the dbDNA system. When reprogramming fibroblasts using protocols employing animal-free constituents, both vectors displayed a much-reduced functional efficiency in terms of iPSC colony production. This was presumably due to a functional advantage conveyed by the inactivated mouse embryonic feeder (iMEF) layer which was absent. Inactivated feeder cells have been evidenced to support targeted cell growth mediated by the release of growth factors (LIF, FGF2 etc), their ability to detoxify culture medium and also to synthesise extracellular matrix proteins (Llames et al., 2015). The supportive nature of feeder cells was further substantiated within the literature. Experiments were conducted analysing the growth of cells in feeder maintained and feeder-free conditions, where the author suggested that, for the above mentioned reasons, feeder maintained growth was more beneficial and supported a more prolonged, undifferentiated growth of pluripotent cells (Richards et al., 2002). Despite functioning at a reduced efficiency in comparison to the feeder produced protocol, the dbDNA system was able to produce viable, bona fide iPSCs when incorporated into the *xenofree* protocol.

A successful protocol for the derivation of iPSCs from both blood and urine-derived cells was produced and recorded within our results. A protocol for the derivation, culturing and reprogramming of urine-derived cells was successfully established within the McKay lab. This was a time-consuming process, but one that offered the prospect of isolating cells for reprogramming in the least invasive manner possible. The dbDNA system was not successful in the derivation of iPSCs from urine-derived cells. The reasoning behind this has yet to be established and was most likely a technical failure. With greater research efforts to improve the protocol and to attempt different vector transfection methods, the system may still be a success within the derivation of iPSCs. That said, a mechanistic failure could still be a possibility that should not be ignored.

There was a plethora of potential mechanistic reasons behind the lack of success on this occasion with the dbDNA system. The doubling times for urine-derived cells was not dissimilar to that of dermal fibroblasts (~24 ± 6 hours and 21 ± 3 hours respectively) and so it seemed unlikely that the issue was in relation to an increased dilution of the dbDNA system due to its lack of tethering capacity (Bolton and Barranco, 1975; Lang et al., 2013). However, upon reprogramming, the urine-derived cells were done so according to a feeder-free protocol; with no supporting iMEFs. As prior mentioned, when reprogramming fibroblasts, the presence of a supporting iMEF feeder layer was a critical element providing the process with a greater reprogramming efficiency with both vector types. Inactive MEFs have been previously demonstrated within the literature to secrete growth factors that support the maintenance of pluripotency within iPSCs and hESCs. When employing a feeder-free protocol, our research group adopted the use of the Matrigel matrix. Matrigel is a soluble basement membrane being extracted from Engelbreth-Holm-Swarm tumours. The membrane was rich in matrix proteins such as laminin, collagen IV, entactin and heparin sulfate proteoglycan (Bissell et al., 1982). Such extracellular matrices (ECM) are important to stem cell growth and survival by recapitulating *in-vivo* stem cell niches and specialised microenvironments (Higuchi et al., 2011; Lee et al., 2011). Moreover, a finite amount of soluble growth factors were present in Matrigel, too. This provided the basis of how Matrigel became increasingly successful and a staple of the urine-derived cell reprogramming protocol. However, iMEF feeder layers can constitutively express growth factors which contribute to the maintenance of pluripotency. Secretion of factors such as *Activin A*, *FGF* (which in-turn activates *PI3K/AKT* pathway demonstrated increased PSC survival), *Wnt* signalling, TGFβ and *NODAL* are all demonstrated to inhibit PSC differentiation are all collectively contributors to successful iPSC development (Beattie et al., 2005; Saha et al., 2008; Romorini et al., 2016; Cristo et al., 2017). In a number of studies it has been demonstrated that inhibiting TGFβ, *NODAL & Activin A* resulted in a proportionate increase in spontaneous differentiation within culturing hESCs (Saha et al., 2008). Likewise, *Wnt* signalling has been implicated in a reduction of differentiation secondary to its increased interactions with *β-catenin* (Nusse, 2008). The secretion of such factors clearly demonstrated an increased ability to maintain PSCs and their pluripotent capacity with great longevity. Duly, the perpetual, constitutive release of factors which

synergistically functioned to maintain a pluripotent state (*Wnt*, TGFβ, *NODAL & Activin)* whilst stimulating cellular proliferation (FGF, EGF etc) by iMEFs provided them with a supporting role in pluripotency induction and maintenance. This continued expression of pluripotent-supporting factors, I believe, was what drove a greater level of reprogramming success on feeder layers in comparison to the feeder-free protocol. The dbDNA system was designed as a transient system with no scaffolding capabilities. Therefore, a dual synergy between the dbDNA vectors transient nature and the lacking feeder layer may potentially be the determining factors in relation to the lack of success on this occasion using the dbDNA system within urine-derived cells. Despite a successful attempt at reprogramming urine-derived cells with the *oriP*-EBNA1 system, the vector was un-successful several times during this process – with AP positive primary colonies being produced only on a single occasion. This helped to substantiate that the failure was not just down to the dbDNA system and that alone, that in fact even the most robust of vectors that have the functional advantage of DNA scaffolding capabilities also struggled with this process.

In the context of reprogramming peripheral blood, the dbDNA system was also unable to induce pluripotency on this occasion too. This again could be for a number of reasons, both technical or mechanistic. I believe, again, the dbDNA system and its reduced success in this instance was attributable to the reduced retention of the vector in comparison to the *oriP*-EBNA1 system. In the protocol for reprogramming peripheral blood, iMEFs were adopted which should support the development of iPSCs at a greater efficiency for both vector types. However, the protocol also incorporated the use of *Dynabeads* to activate CD3/CD28+ cells and therefore increased T-cell proliferation between '*100 and 1000-fold*'. The increased proliferation rate will, most likely, result in an increased dilution of the dbDNA system and not the *oriP*-EBNA1 system due to the vectors inherent tethering capacity. This ultimately prohibited the induction of pluripotency onset within adherent PBMCs using the dbDNA system. The dbDNA vector did, however, seemingly result in the formation of some cells of a different phenotype from the feeder layer, but that do not resemble iPSCs. This was potentially partially reprogrammed cells. There was a possibility that some transfected cells were able to persist and progress partly through the initiation phase of reprogramming but did not fully transition through it - resulting in only a partially reprogrammed cell state (David and Polo, 2014).

Taken in its entirety, the dbDNA system on this occasion did not function to reprogram and produce iPSCs within cell types other than the previously reported fibroblast samples. The failure of the vector to do could be in relation to either technical or mechanistic failings on mine or the vectors behalf. However, just because the vector did not function to produce iPSCs on this occasion does not mean that it never will. The production of iPSCs from fibroblasts using the dbDNA system was an extraordinary event that in theory should not have occurred with the relative ease with which it did. Therefore, I do believe that with further research efforts that the dbDNA system can be successfully incorporated into protocols for the development of iPSCs from other cell types such as urine and blood-derived cells. But, if not, then a greater mechanistic insight will be gained into the vector to understand its reduced functionality in this aspect and how this might be improved. Either way, something will be gained from future experiments undertaken within this remit.

### 3.7. Future work:

In my opinion the future line of work should follow trying to meet relevant criteria in order to truly establish the first clinical grade dbDNA iPSC line. The established criteria involves a host of experiments determining mycoplasma presence, endotoxin, bacteriology/virology testing, SNP testing, Karyotyping, tests for residual vector alongside pluripotency & differentiation tests (Sullivan et al., 2018). Moreover, comprehensive terminal differentiation studies should likewise be undertaken using dbDNA iPSCs to form neurons, cardiac cells, hepatocytes etc to demonstrate the ability of these iPSCs to form specific and potentially therapeutic cell types. An integration analysis should also in my opinion be carried out to determine the presence of vector DNA within the iPSC genomic DNA. A whole genome sequencing or exome sequencing experiment would unequivocally rule out the potential of the dbDNA system to integrate into the iPSC genomic DNA and so would be a future experiment to consider. Moreover, as an internal quality control and to also determine the effects that the *shp53* transgene has on *oriP*-EBNA1 iPSCs in comparison to the dbDNA system, a ddPCR looking at p53 associated mutations in *oriP*-EBNA1 and dbDNA iPSC lines would be of interest. This would help to elucidate any underlying mutations within the dbDNA produced iPSCs and at what rate they occur or are incited at in comparison to the *oriP*-EBNA1 system.

## 3.8. Summary:

The dataset produced by myself during this PhD demonstrated a proof-of-principle that the dbDNA system can be successfully applied to the process of iPSC production from human dermal fibroblasts; the most common source of cells for reprogramming. The iPSCs produced by the dbDNA system adopted a perpetual self-renewal phenotype as well as being prompted to differentiate forming cells of all three germ lineages. The iPSCs produced by the dbDNA system have been evidenced to be, without doubt, *bona fide* iPSCs. The dbDNA system also has the capacity to produce iPSCs when applied to a *xenofree* reprogramming protocol. In the context of reprogramming, the dbDNA system demonstrated a functionality that was competitive in relation to the *oriP*-EBNA1 system, and as the literature states, other non-integrating reprogramming systems too. The extraordinary nature of this result in relation to the Doggybone system however, was that the dbDNA system functioned independent of scaffolding capabilities or additional transgenes which have been evidence to improve reprogramming efficiencies. Following global transcriptome analysis, the dbDNA produced iPSCs exhibited a reduced pro-inflammatory environment, alongside greater maintenance of its pluripotent nature and its genomic integrity in comparison to the *oriP*-EBNA1 system. Taken together, the dbDNA system presents as a vector which has the capacity to produce safer and more stable iPSCs in comparison to its plasmid counterpart; a counterpart which has become well-established as the *gold standard* of iPSC development. Therefore, if we have been able to describe a functional reprogramming system which remedies many secondary issues associated with the EBNA1 plasmid system, there is no reason why, with further investment, the dbDNA system cannot itself become the next pluripotency *gold standard*. I feel that this, in conjunction with my outlined *future work* could result in the beginning of a new era of iPSC development, using clinical grade dbDNA vectors as opposed to ones rooted firmly in the manipulation of bacteria and viruses. The production of a cGMP dbDNA-iPSC line could kickstart the incorporation of dbDNA iPSCs into potential future clinical trials which will exponentially increase as the field of stem cell biology also continues to develop.

# 4.0. Materials and Methods

## 4.1. Cell Culture:

Table 2: Cell culture reagents.

| Reagent | Company | Catalogue Nº |
|---|---|---|
| DMEM | Sigma | D6546 |
| mTeSR™ 1 | Stemcell Technologies | 05851/05852 |
| DMEM/F12 | Gibco, Life Technologies | 31331093 |
| XVIVO10 | Lonza | 04-743Q |
| REBM™ Renal Epithelial Cell Growth Basal Medium | Lonza | CC-3191 |
| OptiMEM | | |
| Dulbecco's Phosphate buffered saline | Sigma | D8662 |
| MEM Non-essential amino acids | Life Technologies | 1140050 |
| Plasmocin | Invivogen | Ant-mpp |
| Foetal Bovine serum | Gibco, Life Technologies | 10270-106 |
| L-Glutamine | Sigma | G7513 |
| Penicillin/Streptomycin | Sigma | P0781 |
| Dulbecco's Phosphate buffered saline | Sigma | D8662 |
| Human Dermal Fibroblast Nucleofector™ Kit | Lonza | VPD-1001 |
| Knockout serum replacement | Gibco, Life Technologies | 10828-028 |
| β-mercaptoethanol | Life Technologies | 31350-010 |
| Mitomycin C | Sigma Aldrich | M4287 |
| TrypLE express enzyme | Gibco, Life Technologies | 12605028 |
| FGF2 | R&D Systems | 233-FB-025 |
| Laminin | Millipore | CC095 |
| Gelatin from porcine skin | Sigma | G1890 |
| Rock Inhibitor (Y-27632) | Sigma | Y0503-1mg |
| Dimethyl sulfoxide | Sigma | BP231-100 |
| Matrigel | BD Bioscience | 734-0270 |
| Rock inhibitor (Y-27632) | Sigma | Y0503-1mg |
| Polyethylenimine | Sigma | **03880** |
| UltraPure™ 0.5M EDTA, pH 8.0 | ThermoFisher Scientific | 15575020 |
| Ficoll® Paque Plus | Sigma | GE17-1440-02 |
| rIL-2 | Peprotech | 200-02 |
| Gibco™ Dynabeads™ Human T-Activator CD3/CD28 | ThermoFisher Scientific | 111.61D |
| REGM™ Renal Epithelial Cell Growth Medium SingleQuots™ Kit | Lonza | CC-4127 |
| PDGF-AB | Peprotech | 100-00AB |
| hEGF | Peprotech | E9644-.2MG |
| SIGMAFAST™ BCIP®/NBT | Sigma | B5655-25TAB |
| Essential 6™ medium | ThermoFisher Scientific | A1516401 |
| Essential 8™ medium | ThermoFisher Scientific | A1517001 |
| Vitronectin | StemCell Technologies | 07180 |

HEK293T cells were acquired from Dr Steve Howe – institute of child health, UCL. Control neonatal dermal fibroblasts (nhDF) were purchased from Fisher Scientific (C0045C) and PromoCell (CAT NO). CLN3 and CLN6 hDFs were obtained from Prof. Sara Mole from the Laboratory for Molecular Cell Biology at UCL as part of the BATcure Horizon2020 consortium. Furthermore, Shef3 Human Embryonic stem cells (hESCs) were procured from the UK stem cell bank (SCSC10-48). Finally, MEF feeder cells were purchased from Cambridge Bioscience (CBA-310).

## 4.2. Cell culture methodologies:

### 4.2.1. The culturing and inactivation of Mouse embryonic fibroblasts (MEFs):

MEFs were cultured in complete DMEM (DMEM (4500mg/L glucose), 10% FBS (v/v), L-Glutamine (4mM), Penicillin/Streptomycin (1x)) supplemented with 1x (v/v) non-essential amino acids. The cell medium was replaced every other day and upon reaching confluency (90-95%) the cells were passaged in a 1:4 ratio using 150µL/$cm^2$ of TrypLE. The cells were centrifugated at 258 G for 5 minutes before being re-suspended in an adequate volume of culturing media and re-plated.

Following amplification to passage 4 (P4), the MEFs were mitotically inactivated. The cells were incubated in complete DMEM with Mitomycin C (0.1µg/uL) at 37°C for 3 hours. Post-incubation, the MEFs underwent a minimum of 3 wash steps in 10mL of Dulbeccos Phosphate Buffered Saline (DPBS) before again being enzymatically detached using TrypLE. Once the cells had dis-associated from the flask surface, they were stored at -80°C in FBS supplemented with 10% (v/v) Dimethyl Sulfoxide (DMSO) at a density of ~5 x $10^6$ cells/mL. Once required, MEFs were defrosted and seeded at a density of 5 x $10^4$ cells/$cm^2$ on culture dishes/flasks that had been pre-coated in 0.1% (w/v) gelatin.

4.2.2. Culturing and passaging of Pluripotent stem cells (PSCs):

Human embryonic stem cells (hESCs) and induced pluripotent stem cells (iPSCs) – both of which are considered PSCs - were cultured on MEF feeder layers (iMEFs) in hESC media. The medium was refreshed every other day. PSC colonies were passaged regularly - every 4-10 days – determined mostly by colony morphology. Upon passaging, fresh hESC media was placed onto the cells. PSC colonies were then manually excised from the flask into the fresh media. The colonies were then further dissociated by being passed through a pipette before being placed onto a fresh iMEF feeder layer.

4.2.3. Human Dermal Fibroblast (HDF) culturing and maintenance:

HDFs were cultured in complete DMEM with regular media changes every other day. The cells were routinely passaged, being dissociated using TrypLE before centrifugation at 258G for 5 minutes. The cells were then seeded at a density of ~3 x $10^4$ $cm^2$.

4.2.3. Somatic cell reprogramming & production of iPSCs:

4.2.3.1. Reprogramming of Human dermal fibroblasts (hDFs):

Episomal plasmid-based reprogramming factors (SOX2, OCT4, KLF4, l-Myc, shp53) were prepared to concentrations outlined in Table 4. Additional EBNA1 expression plasmid was also included to facilitate an improved reprogramming efficiency. Doggybone vectors were prepared as outlined in Table 4. The dbDNA system did not employ the use of the EBNA1 protein or the shp53 protein.

Table 4: Reprogramming plasmids and the concentrations at which each vector is utilised.

| *Construct* | *DNA amount (*µg) | *Catalogue number* |
|---|---|---|
| *pCXLE-hSK* | 2.0 | Addgene ID: 27078 |
| *pCXLE-hUL* | 2.0 | Addgene ID: 27080 |
| *pCXLE-hOCTshp53* | 2.0 | Addgene ID: 27077 |
| *pCXLE-EBNA1* | 2.0 | Addgene ID: 37624 |
| *DbDNA-hSK* | 2.33 | n/a |
| *DbDNA-hUL* | 2.33 | n/a |
| *DbDNA-OCT4* | 2.33 | n/a |

Nucleofection solution, consisting of 90µL NHDF Nucleofector™ solution + 20µL of Supplement 1 (LONZA: VPD-1001) was firstly prepared. 8µg of episomal DNA/dbDNA was then deposited into the solution. ~4.5 x $10^5$ HDFs were trypsinised and centrifuged at 258G before being re-suspended in the Nucleofector/DNA mixture and transferred into a glass cuvette. The cells were then nucleofected (*P-022* programme*: Human dermal fibroblasts – high viability*). The fibroblasts were then seeded onto a single 6-well in complete DMEM – this would be considered day 0.

On day 1, the medium was refreshed before then being replaced every 2 days. Upon reaching confluency (>90%), the nucleofected HDFs could then be passaged and seeded into a T75$cm^2$ flask. On day 8, the re-programming HDFs were dissociated using TrypLE before 6x$10^4$ cells were re-plated onto a T25$cm^2$ flask containing a feeder layer in 5mL complete DMEM. On day 9, the cell medium was exchanged from complete DMEM to hESC media which was likewise replaced every 2 days until colony formation and clonal excision.

4.2.3.2. Reprogramming of fibroblasts using a xenofree protocol:

Human dermal fibroblasts were cultured using Essential 6 media prior to the introduction of reprogramming plasmids via Amaxa nucleofection. The reprogramming protocol initiated and carried out was identical to that described in 4.2.3.1 using HDFs. However, the iMEF feeder layer the reprogramming fibroblasts were introduced onto 7 days post nucleofection was replaced by vitronectin coated wells. Vitronectin was thawed at room temperature before being diluted in DPBS to 10µg/mL (v/v). Wells of a 6-well plate were then coated with vitronectin for a minimum of 1 hour prior to use. Secondly, the traditional hESC medium which the reprogramming cells are normally cultured in on Day 8 onwards for the reprogramming protocol is replaced by *Essential 8*. Once primary colonies are ready to be excised, they are re-plated again using vitronectin and *Essential 8* medium.

4.2.3.3. Reprogramming of Peripheral blood:

Peripheral blood was isolated into an appropriate anti-coagulant containing vacutainer. The sample was then immediately processed being diluted at a minimum of 1:1 (v/v) with DPBS supplemented with 0.5M EDTA. With care not to disturb the interface between the two liquids, 6mL of diluted blood was gently layered atop 3mL of Ficoll. Isopycnic centrifugation (400G for 35 minutes at 18°C) was then undertaken. This was to fractionate the blood and ficoll based on their relative densities. The buffy coat layer (in-between the plasma and ficoll layers) was then carefully extracted before being washed in DPBS twice. 3-5x$10^6$ cells were then resuspended into 110µL of Nucleofection solution with 4µg of episomal DNA/dbDNA. The cells were then nucleofected using the *V 0-24* programme on the Amaxa 2b nucleofector. The product was then seeded onto an iMEF layered 6-well in xVIVO10 medium. This was furthermore supplemented with 30 U/mL rIL-2 and 5uL Dynabeads® for human T-cell activation. 2 days post nucleofection, hESC media was then added without aspirating the previous xVIVO10 medium. 4 days post nucleofection, all the medium was then removed and 2mL hESC was then added and refreshed every other day until colony formation.

4.2.3.4. Urine reprogramming:

Pre-sample collection & collection process:

1 hour prior to sample collection, a participant was advised to drink plenty of water to increase the propensity of cellular shedding and collection. Once ready, the participant was advised to clean the urethral area using an anti-septic wipe prior to any sample collection. The first stream of urine was not collected, the rest was collected in a sterilised beaker.

Post collection/Cellular isolation:

The next steps of the process were carried out in a sterile environment. The sample was transferred from the beaker into 50mL Falcon tubes. The samples were centrifuged at 200g for 10 minutes. During this, 12-well plates should be gelatinised and placed in an incubator at 37°C for ~20 minutes. Once centrifuged, the sample supernatant was discarded from each tube and 1mL of primary medium (DMEM/high glucose and Ham's F12 nutrient mix (1:1), FBS (10% (v/v)), Penicillin/Streptomycin (1x), 1 x REGM™ Renal Epithelial Cell Growth Medium SingleQuots™ Kit (CC-4127)) was added. The primary medium and any cells were transferred to the gelatinised 12 well-plate where an additional 1mL of primary medium was added to each well. The cells were then incubated at 37°C overnight.

Cell Expansion:

24 hours post isolation, 1mL of primary medium was added to each well without removing any of the previous media. This was repeated for the next 3 days. 96 hours after the initial plating of cells, 3 of the 4mL of primary medium was removed. Two different proliferation media were then produced. RE proliferation media (REBM™ Renal Epithelial Cell Growth Basal Medium (CC-3191), 1 x REGM™ Renal Epithelial Cell Growth Medium SingleQuots™ Kit) and MC proliferation media (DMEM w/high glucose, 10% FBS (v/v), 1% GlutaMAX (v/v), 1% NEAA (v/v), Penicillin/Streptomycin (1x), bFGF (5ng/mL), PDGF (5ng/mL), EGF (5ng/mL)) were produced. Then, 1mL RE/MC proliferation media was added to each well. From this point onward, half of the media was replaced with RE/MC proliferation media daily. Upon reaching confluency, the cells should be passaged using TrypLE.

Reprogramming:

Upon reprogramming, between 5 x $10^5$ – 1 x $10^6$ cells are required. A total of 6µg of plasmid/dbDNA (Table 5) was nucleofected using the Amaxa 2b Nucleofector with the programme *T-020*. The cells were subsequently plated onto Matrigel coated wells in MTESR medium. The medium was refreshed daily until colony formation.

Table 5: Vector concentrations utilised in the reprogramming of urine derived cells.

| Construct | DNA amount (µg) |
|---|---|
| hSK (SOX2 + KLF4) | 1.75 |
| OCT/shp53 | 1.75 |
| hUL (LIN28 + L-Myc) | 2.5 |
| Db-hSK | 1.75 |
| Db-hOCT | 1.75 |
| Db-hUL | 2.5 |

### 4.2.4. Colony counting & reprogramming efficiency:

The efficiency of reprogramming was calculated to determine the ability of each vector type to induce the production of iPSCs. This was calculated as a percentage of the number of colonies produced by each vector type relative to the total number of fibroblasts re-plated onto the iMEF feeder layer on day 8 of the protocol.

On day 21, alkaline phosphatase (AP) staining was carried out to determine the presence of fully reprogrammed iPSCs. This would stain the legitimate iPSC colonies; whose expression of AP is much higher than the feeder layer and most other cell types. These colonies could then be counted to determine an efficiency.

### 4.2.5. HEK293T maintenance and transient transfection:

HEK293T cells were seeded at 2-5 x $10^4/cm^2$ in complete DMEM medium. The medium was replaced with fresh media every 2 days. Once a confluency of 80-90% was reached, the cells were passaged using TrypLE.

For a transient transfection, HEK293T cells were cultured to a confluency of ≥90%. Equal volumes (Table 6) of Polyethylenimine (PEI) and DNA were aliquoted into

separate tubes containing OptiMEM. The tubes containing PEI and DNA were then combined and mixed before a 20-minute incubation step at RT. During this incubation, 300µL of OptiMEM was also added to the well of HEK293T cells to be transfected. Afterwards, the PEI-DNA complexes were added to the cells before a 2-3-hour incubation at 37°C followed. Afterwards, any un-transfected or remaining PEI/DNA complexes were removed from the HEK293T cells before the addition of 2mL of complete DMEM.

Table 6: Preparations of PEI and DNA for the transfection of a vector into cells.

**PEI preparation:**

| Optimem Volume (µL) | PEI Volume (µL) |
|---|---|
| 327.25 | 2.75 |

**DNA Preparation:**

| Optimem Volume (µL) | DNA amount (µg) |
|---|---|
| 327.25 | 2.75 |

## 4.3. Molecular Biology:

Table 3: Molecular biology reagents.

| Reagent | Company | Catalogue No |
|---|---|---|
| RIPA buffer | Thermo Scientific | 89900 |
| Bradford reagent | BioRad | 500-0006 |
| Acrylamide/bis-acrylamide 40% | Sigma | A7802 |
| TEMED | Sigma | T9281 |
| Tween 20 | Sigma | 19379 |
| Precision Plus Protein™ Kaleidoscope™ Pre-stained Protein Standards | Biorad | 1610375 |
| Marvel original dried skimmed milk | Supermarket | |
| Protease inhibitor cocktail | Sigma Aldrich | P8340 |
| Phosphate buffered saline tablets | Sigma | P4417 |
| Immobilon Western Chemiluminescent HRP Substrate | Merck | WBKLS0500 |
| APS | Sigma | 215589-100G |
| Blot absorbent filter paper | Biorad | 1703932 |
| Methanol | Fisher scientific | |
| Glycine | Sigma | G8898-500G |
| Tris base | Fisher scientific | BP152-1 |
| Sodium dodecyl sulphate | | |
| Proteinase K | Fisher scientific | 10172903 |
| Qiaprep spin miniprep kit | Qiagen | 27106 |
| RNeasy mini kit | Qiagen | 74104 |
| RQ1 RNase-free DNase kit | Promega | M6101 |
| dNTPs | Promega | PRU1240 |
| M-MLV reverse transcriptase | Promega | M1701 |
| RNasin plus inhibitor | Promega | N2611 |
| Random primers | Promega | C1181 |
| KAPA SYBR FAST universal 2x qPCR master mix | KAPA Biosystems | KK4601 |
| GelRed nucleic acid stain | VWR International | 730-2958 |
| Agarose (NMA) | Sigma | A5093 |
| O'Generuler ladder mix | Fisher Scientific | 1188393 |
| Isopropanol | Sigma | 190764 |
| 4% Paraformaldehyde | | |
| Triton 100x | Sigma | T8532 |
| Bovine serum albumin | Sigma | 05479 |
| PVDF Transfer Membranes | Thermo-Fisher | 22860 |
| NucBlue™ Live ReadyProbes™ Reagent | Thermo-Fisher | R37605 |
| UltraPure™ 0.5M EDTA, pH 8.0 | Thermo-Fisher | 15575020 |

| | | |
|---|---|---|
| Q5® High-Fidelity DNA Polymerase | New England BioLabs | M0491L |
| ISOLATE II Genomic DNA Kit | Bioline | BIO-52066 |
| UltraPure™ Low Melting Point Agarose | Thermo-Fisher | 16520050 |
| Sodium Chloride | Merck | S3014 |
| Sodium Hydroxide | Merck | 795429-500G |
| Dimethyl sulfoxide | Merck | BP231-100 |
| Anhydrous sodium acetate | Merck | W302406 |
| SYBR™ Gold Nucleic Acid Gel Stain (10,000X Concentrate in DMSO) | Thermo-Fisher | S11494 |
| GenElute™ Gel Extraction Kit | Merck | NA1111-1KT |
| North2South™ Chemiluminescent Hybridization and Detection Kit | Thermo-Fisher | 17097 |
| Nylon membranes | Thermo-Fisher | LC2003 |
| EcoRI | New England BioLabs | R0101S |
| Ficoll® 400 | Merck | F2637-5G |
| Polyvinylpyrrolidone (PVP) | Merck | P5288-100G |

### 4.3.1. Protein quantification & Western Blotting analysis:

#### 4.3.1.1. Protein quantification

Cells were lysed for protein isolation utilising RIPA buffer augmented with 1:1000µL protease inhibitor cocktail. The lysates were then stored at -80°C until required.

Each lysate underwent quantification using the Bradford assay. Bradford reagent was diluted at a ratio of 1:5 (v/v) with water. 195µL diluted Bradford reagent was then added in triplicate to wells of a 96-well plate. 5µL Protein standards were added to wells containing diluted Bradford with concentrations ranging from 0.5µg - 8µg per 5µL. Experimental lysates were diluted 1:10 (v/v), thereby minimizing the effects of RIPA on the Bradford reagent. 5µL of each diluted lysate was then added in triplicate. The wells were mixed prior to spectrophotometric quantification at 630nm.

#### 4.3.1.2. SDS-PAGE gel production:

Following protein quantification using the Bradford assay, the constituents of a bis-acrylamide separating gel were added to a 50mL Falcon tube. The contents were then carefully decanted into a glass cast. Isopropanol was then carefully added to the top. This was to ensure a uniform interface between the separating gel and stacking gel whilst also preventing any oxygen-dependant interference with gel polymerisation. Upon solidification of the separating gel, the isopropanol was removed. A stacking gel was then subsequently made before being poured on top of the separating gel and a comb inserted. The gel was left to solidify.

Table 7: Recipe to produce the stacking and separation gels for western blotting.

Separation gel:

| Reagent | Volume |
|---|---|
| Solution A (1.5M Tris pH8.8 / 0.4% (w/v) SDS) | 2.5mL |
| 30% Bis/acrylamide | 5mL |
| Water | 2.5mL |
| TEMED | 10µL |

Stacking gel:

| Reagent | Volume |
|---|---|
| Solution B (0.5M Tris-HCl pH6.8 / 0.4% (w/v) SDS) | 2.5mL |
| 30% Bis/acrylamide | 5mL |
| Water | 2.5mL |
| TEMED | 10µL |
| APS (10%) | 30µL |

#### 4.3.1.3. Electrophoresis & Protein transfer:

The SDS-PAGE gels were submerged in running buffer (Tris base (25mM), 0.1% SDS (w/v), Glycine (190mM), 1L dd$H_2O$) within the tank before 15-30µg of each protein sample was loaded into wells alongside a single well containing 5µL Kaleidoscope protein ladder. The gel was then electrophoresed for 105 minutes at 90v. After sufficient migration, the separated protein lysate was then transferred from the SDS-PAGE gel onto a PVDF membrane. The gel containing the separated protein was placed atop the PVDF membrane. Both were then sandwiched between blotting paper which had been soaked in blotting buffer (Tris Base (25mM), 20% Methanol (v/v), Glycine (190mM), 1L dd$H_2O$). The sandwiched membrane was furthermore placed into the semi-dry blotter and transferred at 15v for a minimum of 60 minutes - permitting complete transfer of protein from the gel to the membrane via capillary action.

4.3.1.4. Primary antibody addition:

Following protein transfer, the membrane was carefully removed using sterilized tweezers before being blocked in DPBS supplemented with 5% (w/v) milk and 0.1% (v/v) Tween20 for a minimum of 60 minutes. After blocking, primary antibodies were diluted in blocking buffer before being placed atop the membrane. This was then incubated on a rocker at 4°C overnight.

Table 8: Antibodies used for western blotting.

| Antibody | Dilution | Catalogue N$^{o}$ |
|---|---|---|
| OCT3/4 | 1:400 | Abcam (Ab18976) |
| SOX2 | 1:2000 | Biotechne (AF2018) |
| LIN28 | 1:1000 | R&D (AF3757) |
| β-actin | 1:10,000 | Sigma Aldrich (A2228) |

4.3.1.5. Secondary antibody detection:

Following incubation with a primary antibody, the membrane was washed with PBS with 0.1% (v/v) tween20 for 5 minutes (per wash). A HRP-conjugated secondary antibody was then added to blocking buffer at a 1 in 2000 dilution. This was added to the membrane on a rocker at room temperature for a minimum of 60 minutes. The membrane was washed again prior to the addition of a chemiluminescent HRP substrate detection reagent. This was incubated on the membrane for 1 minute. The blot was then visualized on a transilluminator.

### 4.3.2. Immunocytochemistry (ICC):

Culture medium was firstly removed from cells to be analysed. The cells were then washed with DPBS (x3). PBS supplemented with 4% paraformaldehyde (PFA) (v/v) was used to fix the cells at room temperature for 20 minutes. The cells were washed following fixation, with DPBS (x3). The wells to be stained were then permeabilised if the protein of interest was not membranous. Permeabilization was carried out utilising 0.3% Triton X (v/v) in PBS for 10 minutes at room temperature. Further DPBS washes (x3) were carried out before the cells were blocked for a minimum of 30 minutes. Blocking buffer was produced using PBS supplemented with 2% bovine serum albumin (BSA) (w/v) + 0.1% (v/v) Tween20. Primary antibodies were then diluted to appropriate concentrations (table) in blocking buffer and were added to the cells before an incubation at 4°C overnight. The cells were thereafter washed before the secondary antibody was diluted (1:500) in blocking buffer and left on the cells for 1 hour at room temperature in the dark. Following further wash steps, DAPI was added in PBS for 1 minute before being removed and the cells visualised on a fluorescent microscope.

Table 9: Antibodies used for ICC analysis.

| Antibody | Dilution | Catalogue Nº |
|---|---|---|
| OCT4 | 1:100 | Abcam (Ab18976) |
| SOX2 | 1:200 | Biotechne (AF2018) |
| Tra-1-60 | 1:200 | Abcam (Ab16288) |
| Tra-1-81 | 1:200 | Abcam (Ab16289) |
| βIII-tubulin | 1:200 | R&D systems (MAB1195) |
| α-Smooth muscle actin (SMA) | 1:100 | Abcam (ab5694) |
| SOX17 | 1:60 | R&D systems (AF1924) |
| H2AX | 1:1000 | Abcam (ab11174) |

### 4.3.3. Reporter gene expression kinetics via FACS analysis:

Cells were firstly trypsinised and counted using a haemocytometer. Then, approximately 4.5 x $10^5$ cells were centrifuged at 258 G for 5 minutes. GFP-expressing plasmid DNA/dbDNA (8µg) was then nucleofected - as per the previously mentioned protocol. Following nucleofection, the cells were then added to 1mL of complete DMEM before being distributed evenly for different timepoints. Upon fluorescence quantification, the cells were trypsinised and pelleted as before (258G, 5 minutes). This was re-suspended in 1mL of DPBS before being transferred into a Falcon® 5mL round bottom polystyrene test tube.

Upon analysis, a negative control of un-transfected HDFs was firstly utilised to gate and quantify any cellular autofluorescence using FITC-A expression (P1). Subsequently, FITC expression beyond the autofluorescence of the control HDFs (P1), was likewise gated as GFP expressing cells (P2). The percentage of cells outside of the P1 gate and within the P2 gate were continually monitored over a 21-day period and quantified as a percentage of 10,000 cells.

### 4.3.4 RNA assays

#### 4.3.4.1. RNA extraction from human cell lines:

Total RNA extraction was carried out utilising the *Qiagen RNeasy Minikit*. The cells to be lysed were collected or re-suspended in the lysis buffer (RLT). Then, for cells that may be difficult to lyse, homogenisation steps were taken to maximise the potential RNA yield. A plastic pestle was used as well as passing the sample through a fine needle. During extraction, the RNA undergoes a DNase treatment on-column using *RQ1 DNase*. *The Minikit* adopts the use of a silica membrane, which binds RNA from cell lysates. The high purity RNA was then eluted from the column using 30µL of RNase free water. The concentration of RNA isolated was then determined using a spectrophotometer (Nanodrop, U.K). The RNA samples were subsequently stored at -80°C.

Prior to use, the quality of RNA was determined by gel electrophoresis. 1µg of RNA was loaded onto a 1.5% gel prepared using 1x TAE (Tris Base (40mM), glacial acetic

acid (40mM), EDTA (1mM)). The gel was submerged in 1x TAE and electrophoresed. If both the 18s and 28s ribosomal bands were clearly observed, it was concluded that the RNA was intact and suitable for any further analysis.

4.3.4.2. cDNA generation from RNA:

A 1µg RNA starting product was mixed with 0.5µg DNA hexamers of randomized sequences for first strand Complimentary DNA (cDNA) synthesis. The RNA/Random primer mixture was then heated at 70°C for 5 minutes – melting template (RNA) secondary structure formation. The sample was then stored immediately on ice to prevent any secondary structure re-formation. The sample was then briefly centrifuged to coalesce the solution to a pool at the bottom of the tube. To the annealed template/primer mixture, the subsequent components were added, in order, to commence reverse transcription. Firstly, 5µL M-MLV reaction buffer (5x), 1.25µL dATP (10mM), 1.25µL dGTP (10mM), 1.25µL dCTP (10mM), 1.25µL dTTP (10mM), 0.6µL (25 units) Recombinant RNasin® Ribonuclease Inhibitor, 1µL (200 units) M-MLV RT, $ddH_2O$ up to 25µL. The sample was then gently mixed before being incubated for 1 hour at 37°C. The reaction was then terminated by heating at 70°C for 10 minutes. cDNA samples were then stored at -20°C until use.

4.3.4.3. Quantitative Polymerase Chain Reaction:

Transcript analysis was performed by qPCR using the KAPA SYBR® FAST qPCR kit in a 96 well-plate format. Each well contained a 10µL total reaction, of this, 5µL of SYBR, 200nM of both forward and reverse primers (0.4µL), 3.6µL $ddH_2O$ and 1µL cDNA was used. A no template control and a no RT control were also incorporated, with $ddH_2O$ replacing the missing component. The qPCR was then carried out on a StepOnePlus (Thermo, U.K). Initial denaturation was carried out at 95°C for 5 minutes, followed by 40 cycles of 95°C for 30 seconds, 60°C for 30 seconds and 72°C for 30 seconds. Primers were often designed with an annealing temperature of 60°C for continuity and to span an exon to prevent amplification of genomic DNA. Relative expression of the analysed genes was normalised to the PABPC4 housekeeping gene and calculated via the $\Delta\Delta$Ct method.

4.3.4.4. iPSC pluripotent RT-PCR characterization:

Endogenous expression of key pluripotency factors was determined through Reverse transcriptase-PCR (RT-PCR) amplification. RNA isolation and subsequent cDNA generation was carried out as prior mentioned. cDNA was then utilised as the starting product for the reaction. PCR was then carried out using Q5 high fidelity DNA polymerase. Each reaction contained: 5µL Q5 reaction buffer (5X), 0.5µL dNTPs (10mM), 2.5µL forward & reverse primer mixture (10µM), 0.25µL Q5 DNA polymerase, 1µL cDNA (40ng), dd$H_2O$ up to 25µL. The samples were briefly mixed before beginning the reaction on the Agilent Sure Cycler 8800 thermocycler. The reaction began with an initial denaturation step of 98°C for 5 minutes. The samples were then subjected to 35 cycles of 98°C for 30 seconds, 55°C for 30 seconds and 72°C for 1 minute. A final extension step of 72°C was carried out for 7 minutes prior to incubation at 4°C. The samples were then analysed by being run on a 1.5% agarose gel produced using TAE and supplemented with gel red (10,000X). The gel was submerged in TAE before the PCR samples were loaded and electrophoresed. The gel was subsequently imaged using a transilluminator.

Table 10: Primers and their forward and reverse sequences used for the pluripotency RT-PCR.

| Target | Forward primer | Reverse primer |
|---|---|---|
| Endogenous OCT4 | GCGATCAAGCAGCGACT | TTCACCTTCCCTCCAACC |
| Endogenous SOX2 | CATGTCCCAGCACTACCAGA | GGGTTTTCTCCATGCTGTTT |
| Endogenous LIN28 | TGTCCAAATGCAAGTGAG | GCAGGTTGTAGGGTGATTCC |
| Nanog | TTTGTGGGCCTGAAGAAAACT | AGGGCTGTCCTGAATAAGCAG |
| E-Cadherin | TGCCCAGAAAATGAAAAAG | GTGTATGTGGCAATGCGTTC |
| RN18S1 | ACACGGACAGGATTGACAGA | GGACATCTAAGGGCATCACAG |

## 4.3.5. DNA-based PCR assays:

### 4.3.5.1. Vector integration:

Genomic DNA was isolated using the ISOLATE II genomic DNA kit. Once prepped, the sample was then used as a starting product to determine potential integrations by reprogramming vectors in clonal iPSC lines. A typical reaction using the Q5 high fidelity DNA polymerase system is outlined in Table 11.

Table 11: recipe for a typical PCR reaction using Q5 DNA polymerase.

| Reagent | Volume (µL) |
|---|---|
| Q5 high fidelity DNA polymerase | 0.5 |
| Q5 5x Reaction buffer | 5 |
| 10mM dNTPs | 0.5 |
| 10µM Forward primer | 1.25 |
| 10µM Reverse Primer | 1.25 |
| Sample | 2 |
| $H_2O$ | 9.5 |

PCR parameters would vary between primer sets. Firstly, a 95°C denaturation step was typically utilised. The primer annealing temperature varied between reactions a gradient PCR was employed to determine the optimum annealing temperature, often between 50-60°C. Finally, an extension temperature of 72°C was used. Q5 DNA polymerase functions at an extension rate of 1kb every 30 seconds.

## 4.3.6. Comet Assay:

### 4.3.6.1. Slide and buffer preparation:

At least 24 hours prior to the start of any comet assay, slides were pre-coated in a 1% (w/v) normal-melting agarose (NMA). This was produced using TAE and kept in warm water to ensure the agarose did not re-solidify. 175µL of agarose was then dotted onto a single slide in duplicate, before coverslips were placed on top. The slides were left

to dry at room temperature for 30 minutes. Once complete, the coverslips were then removed, and the slides allowed to dry overnight at room temperature.

4.3.6.2. Sample loading, lysis & electrophoresis:

Firstly, a 0.7% (w/v) low-melting agarose (LMA) mixture was produced in TAE, boiled and kept in a warm water bath to maintain its liquid state. Any medium was removed from cells to be analysed before being washed in PBS and re-suspended at a concentration of 1 x $10^6$/mL. 60µL of cells (~6 x $10^4$ cells) was then added to 190µL of PBS in an Eppendorf tube. 1mL of LMA was added to the cells before being mixed using a vortex. Immediately after, 75µL of the cell suspension was then placed atop each gel on the pre-made slides (in duplicate) and covered with a 22x22mm glass coverslip. These slides were then placed at 4°C for 5 minutes to set.

Once set, the coverslips were removed from the samples. 190mL of ice cold ready to use lysis buffer (2.5M NaCl, 100mM EDTA, 0.3M NaOH. 5.5% DMSO, 1.1% Triton 100X pH 10) was then decanted into a black slide trough. The slides were left overnight at 4°C.

The following day, the slides were removed and washed for 5 minutes at room temperature in $dH_2O$. Ice cold electrophoresis buffer (3M anhydrous sodium acetate, 10M Tris, pH 9) was then decanted into an electrophoresis tank at 4°C. The slides were subsequently transferred to the electrophoresis tank and left for 20 minutes to equilibrate. A charge of 25v, 300mA was applied for 25 minutes to the slides. Following this, the slides underwent 2 washes in ready to use neutralisation buffer (0.4M Tris Base pH 7.5, diluted to 30%) before being transferred into a dark box to dry overnight.

4.3.6.3. DNA staining and visualisation:

The following day, the slides were firstly rehydrated in $dH_2O$ for 30 minutes. SYBR gold was then diluted to 1x (1:10,000) in $dH_2O$. Once rehydrated, the slides were left to dry for 5 minutes before 100µL of diluted SYBR gold was added to each slide and covered with a 24x60mm coverslip. The slides were then left to stain in the dark for 30 minutes. Once complete, the coverslips were removed, and the slides washed in $dH_2O$ before being dried again for 1 hour. The slides were then ready to image at 10/20x using a fluorescence microscope.

4.3.6.4. Comet analysis:

Analysis of the images was carried out using CaspLab (1.2.3beta2). A minimum of 100 comets were analysed per gel with each technical repeat having 2 gels per slide. 3 biological repeats were analysed per vector.

## 4.3.7. Flow cytometric analysis of cell cycle with propidium iodide: analysis:

Cells were analysed using a Becton Dickinson (BD) FACScalibur flow cytometer (BD bioscience). The software, BD CellQuest Pro™ was utilised to generate a histogram plot. Before analysis, the single cell population was gated using the forward scatter (FS) and side scatter (SS) plot. The propidium iodide histogram was generated by plotting FL-2 versus cell count. Data shown represents 1x $10^4$ cells from the total gated population.

Propidium iodide is a stoichiometric fluorescent intercalating agent that binds DNA in proportion to the amount that is present in the cell. Cells that are in G2 phase and about to undergo cell division contain twice the amount of DNA as cells that are in a quiescent or resting state (G0-G1) so should have an increased fluorescent intensity, detected quantitatively by flow cytometry. Doggybone and oriP-EBNA1 produced iPSCs were expanded before being isolated and mechanically into single cells via pipetting. The cells underwent centrifugation at 300g for 2 minutes. The cells were then re-suspended in DPBS before 2ml of ice-cold 70% ethanol was added dropwise while vortexing to prevent clumping. The cells were left to fix for 30 min at 4°C. After fixation, the culture was centrifuged and washed once with PBS. Again, the supernatant was discarded and 100µl of a 100µg/ml stock solution of RNase A was added to the pellet. 400µl of a 50µg/ml stock of propidium iodide was added and the reaction was kept in the dark for ten minutes.

## 4.3.8. Microarray & unbiased analysis:

### 4.3.8.1. Analysis of raw data:

Microarray analysis was carried kindly carried out at DKFZ German Cancer research Centre, Heidelberg, Germany. The Illumina HT12 beadchip system was used to analyse global gene expression in the samples. RNA was isolated from each of the biological triplicates used in the experiment before being DNase treated. This was a direct hybridization microarray, whereby 750ng of biotinylated cRNA isolated from the different experimental conditions was applied to the chip and hybridized for 17 hours at 58°C. QC analysis was also undertaken by our collaborators in Germany who carried out a quantile normalization using R studio and *the normalize.quantiles* function from the Bioconductor package *preprocessCore.* Following QC analysis, probes from dbDNA & oriP-EBNA1 iPSCs were as such subjected to a student's t-test analysis with a cut off p-value of ≤0.05. From this, the results were then subjected to a Benjamini-Hochberg analysis to determine the False discovery rate (FDR). This was to reduce the possibility of a type 1 error and thereby limit the inclusion of false positive results within the dataset. Again, an FDR cut off ≤0.05 was utilised. Subsequently, fold change expression was calculated between the dbDNA and oriP-EBNA1 samples - probes with a fold change difference of ≥1.5 were taken forward.

Using these significant probes, several different software was used to determine transcription factor enrichment terms. The Reactome database was utilised to project significant probes onto the human genome to help elucidate interacting pathways in relation to cell cycle, metabolism, immune function etc. Moreover, enrichment analysis was likewise undertaken using the MSigDB function from Gene Set Enrichment Analysis (GSEA). This function can then provide information on hallmark genes which summarize and represent specific well-defined biological processes generated via overlaps between gene sets within the MSigDB system. Subsequently, the analysis can provide a p-value for each hallmark process which demonstrates a measure of how significant the changes were for each given gene set – the higher the absolute value of the statistic, the greater its significance. GSEA also provides k/K values, whereby k = the number of genes in the query set and K = the number of genes in the MSigDB database. This can therefore provide information on the direction of change

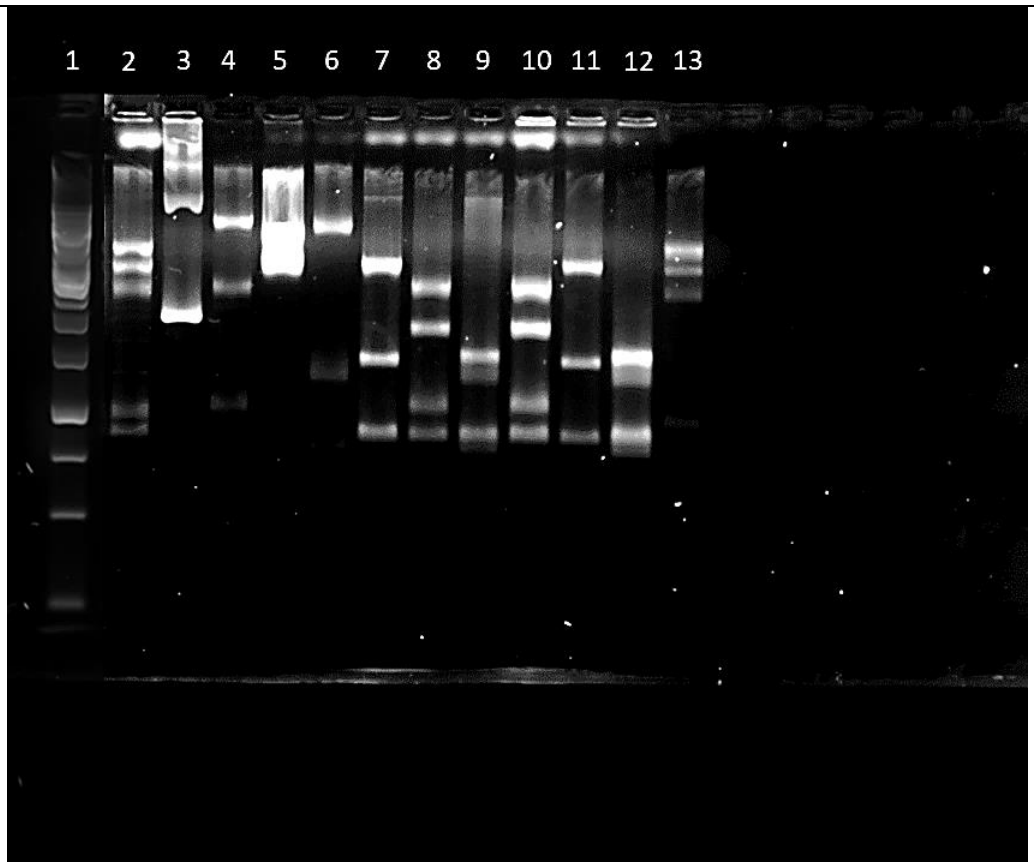
for the biological processes for each significant probe. Finally, a q-value is provided which is an FDR analogue of the p-value after correction for multiple hypothesis testing and again reduces the possibility of including false positive results.

#### 4.3.8.2. Graphics produced using R studio:

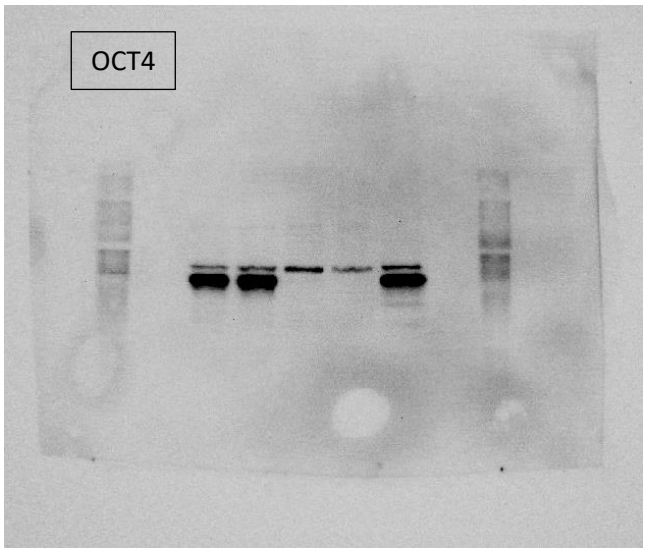
Probe expression values were visualised using images generated from different analyses using R studio. The coding software was used to firstly take the top 1000 highly expressed genes between all data sets and their expression trend compared using a scatterplot. This was further substantiated by the generation of a correlogram for the whole genome of each cell type. A Pearson Product-moment Correlation Coefficient value was generated for each dataset, putting an exact numerical value on the trends between the gene expression of each cell type. Moreover, a Principle component analysis (PCA) was also carried out. This analysis results in the expression of whole transcriptomic datasets for each cell type on a 2D graphic. The distance between the co-ordinates of each plot was therefore proportionate to the respective variability of each cell types transcriptomic profile. Moreover, a volcano plot was also generated permitting the visualisation of DEGs (differentially expressed genes) on a whole transcriptome level. Finally, heatmaps were also generated to permit and simplify the visualisation of large transcriptomic expression profiles between different datasets.

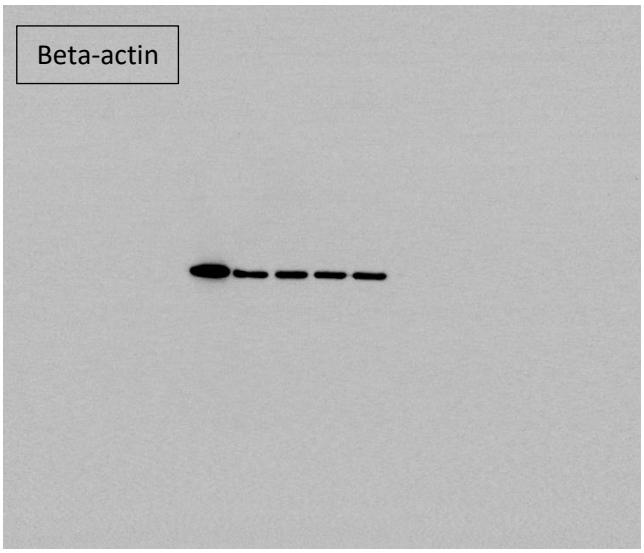
Appendix:

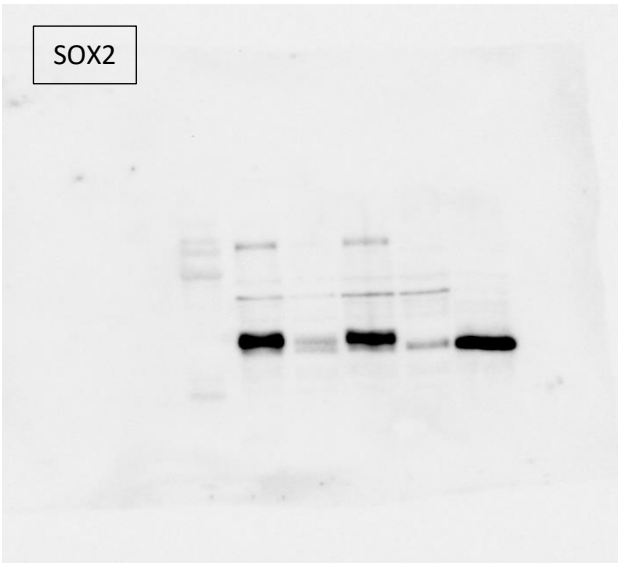
Full image of agarose gel following vector digestion:

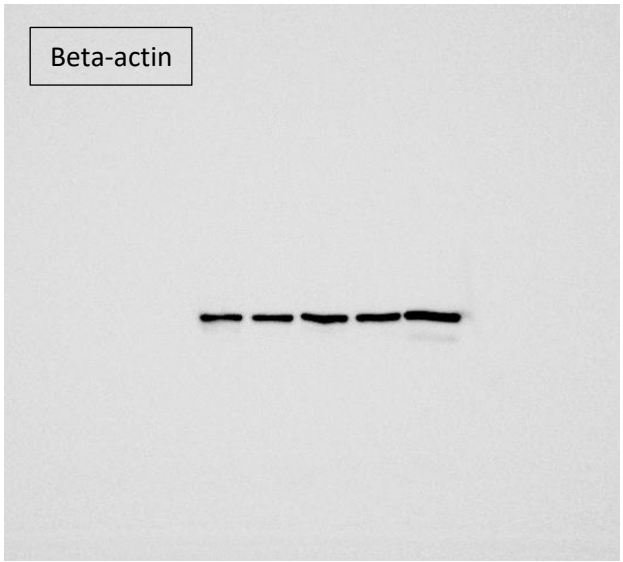


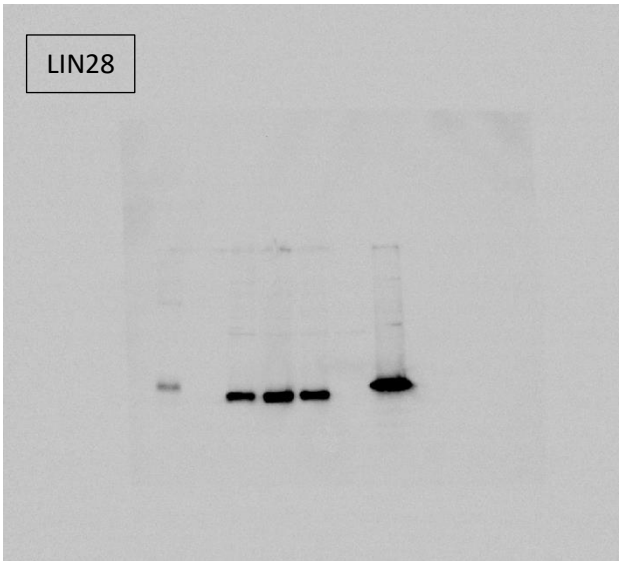
Western blots:

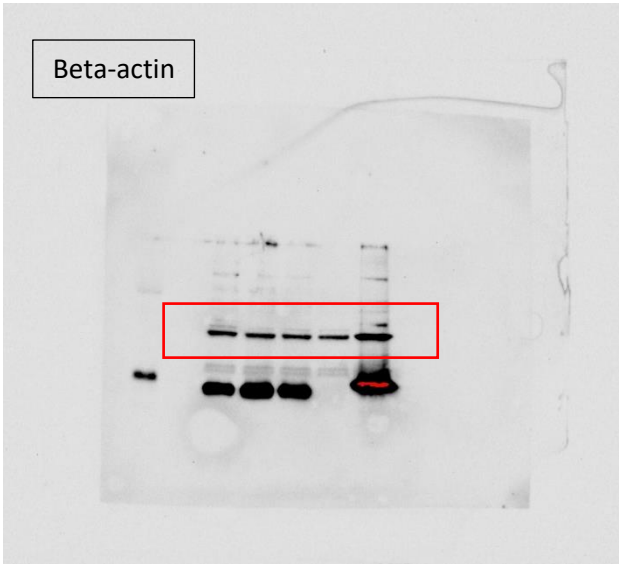



LIN28


Beta-actin

R Script:

Microarray data processing:

Quantile normalisation was carried out with R using the function *normalize.quantities* from the Bioconductor package "*preprocessCore*".

Heatmap script:

```
> heatmap.2(ebna_matrix, Rowv = NA, col = redgreen(75), scale = "row", mar
gins = c(5,10), trace = "none", dendrogram = "column")
> col<-colorRampPalette(redgreen(75))
> heatmap.2(ebna_matrix, Rowv = NA, col = col, scale = "row", margins = c(
5,10), trace = "none", dendrogram = "column")
```

Scree plot and PCA R script:

```
pca <- prcomp(t(Chup), scale=TRUE)


## plot pc1 and pc2
plot(pca$x[,1], pca$x[,2])


## make a scree plot
pca.var <- pca$sdev^2
pca.var.per <- round(pca.var/sum(pca.var)*100, 1)


barplot(pca.var.per, main="Scree Plot", xlab="Principal Component",
ylab="Percent Variation")



pca.data <- data.frame(Sample=rownames(pca$x),
  X=pca$x[,1],
  Y=pca$x[,2])
pca.data


ggplot(data=pca.data, aes(x=X, y=Y, label=Sample)) +
  geom_text() +
  xlab(paste("PC1 - ", pca.var.per[1], "%", sep="")) +
  ylab(paste("PC2 - ", pca.var.per[2], "%", sep="")) +
  theme_bw() +
  ggtitle("My PCA Graph")
```

Volcano plot script:

```
with(topT, plot(lfc, -log10(padj), pch=20, main="Volcano plot - dbDNA - or
iP-EBNA1", cex=1.0, xlab=bquote(~Log[2]~fold~change), ylab=bquote(~-log[10
]~P~value)))
> abline(v=0, col="black", lty=3, lwd=1.0)
> abline(v=-1.0, col="black", lty=4, lwd=2.0)
> abline(v=1.0, col="black", lty=4, lwd=2.0)
```

```
> abline(h=-log10(max(p_Val[padj<0.05], na.rm=TRUE)), col="black", lty=4,
lwd=2.0)
> points(res[2:674,18], -log10(res[2:674,13]), pch=21, bg="green")
> points(res[30197:30662,18], -log10(res[30197:30662,13]), pch=21, bg="red
")
```

Correlogram:

Correlogram was produced using the *GGally* package in R (e.g: corrgram(x, order = , panel=, lower.panel=, upper.panel=, text.panel=, diag.panel=)).

Ethical approval for use of human tissue:

15/08/2018

**Project Title:** Applications of dbDNA in stem cell research

**EthOS Reference Number:** 0536

**Ethical Opinion**

Dear Christopher David Thornton,

The above application was reviewed by the Science and Engineering Research Ethics and Governance Committee and, on the 15/08/2018, was given a favourable ethical opinion. The approval is in place for four years following the

proposed in your application.

**Conditions of favourable ethical opinion**

Application Documents

| Document Type | File Name | Date | Version |
|---|---|---|---|
| Consent Form | Consent form - Ethos | 30/03/2018 | v 1.0 |
| Project Proposal | Project proposal | 04/04/2018 | v 1.0 |
| Information Sheet | Participant info sheet | 24/07/2018 | 2.0 |
| Information Sheet | Advertisement - Reprogramming postter Christopher Thornton | 24/07/2018 | v 1.0 |

The Science and Engineering Research Ethics and Governance Committee favourable ethical opinion is granted with the following conditions

Adherence to Manchester Metropolitan University's Policies and procedures

This ethical approval is conditional on adherence to Manchester Metropolitan University's Policies, Procedures, guidance and Standard Operating procedures. These can be found on the Manchester Metropolitan University Research Ethics and Governance webpages.

Amendments

If you wish to make a change to this approved application, you will be required to submit an amendment. Please visit the Manchester Metropolitan University Research Ethics and Governance webpages or contact your Faculty research officer for advice around how to do this.

We wish you every success with your project.

Science and Engineering Research Ethics and Governance Committee

## 5.0. References:


A, H. and B, S. (2011) ‘Potency of Various Types of Stem Cells and their Transplantation.’ *Journal of Stem Cell Research & Therapy*, 01, January.

Akabayashi, A., Nakazawa, E. and Jecker, N. (2018) ‘The world’s first clinical trial for an aplastic anemia patient with thrombocytopenia administering platelets generated from autologous iPS cells.’ *International Journal of Hematology*, 109, December.

Allen, A., Wang, C., Caproni, L. J., Sugiyarto, G., Harden, E., Douglas, L. R., Duriez, P. J., Karbowniczek, K., Extance, J., Rothwell, P. J., Orefo, I., Tite, J. P., Stevenson, F. K., Ottensmeier, C. H. and Savelyeva, N. (2018) ‘Linear doggybone DNA vaccine induces similar immunological responses to conventional plasmid DNA independently of immune recognition by TLR9 in a pre-clinical model.’ *Cancer immunology, immunotherapy : CII*. 2018/01/12, Springer Berlin Heidelberg, 67(4) pp. 627–638.

Angel, M. and Yanik, M. F. (2010) ‘Innate immune suppression enables frequent transfection with RNA encoding reprogramming proteins.’ *PloS one*. Public Library of Science, 5(7) pp. e11756–e11756.

Anokye-Danso, F., Snitow, M. and Morrisey, E. E. (2012) ‘How microRNAs facilitate reprogramming to pluripotency.’ *Journal of Cell Science*, 125(18) pp. 4179 LP – 4787.

Atkins, J. F., Wills, N. M., Loughran, G., Wu, C.-Y., Parsawar, K., Ryan, M. D., Wang, C.-H. and Nelson, C. C. (2007) ‘A case for “StopGo”: reprogramming translation to augment codon meaning of GGN by promoting unconventional termination (Stop) after addition of glycine and then allowing continued translation (Go).’ *RNA (New York, N.Y.)*. 2007/04/24, Cold Spring Harbor Laboratory Press, 13(6) pp. 803–810.

Babos, K. N., Galloway, K. E., Kisler, K., Zitting, M., Li, Y., Shi, Y., Quintino, B., Chow, R. H., Zlokovic, B. V and Ichida, J. K. (2019) ‘Mitigating Antagonism between Transcription and Proliferation Allows Near-Deterministic Cellular Reprogramming.’ *Cell Stem Cell*. Elsevier, October.

Baghbaderani, B. A., Tian, X., Neo, B. H., Burkall, A., Dimezzo, T., Sierra, G., Zeng, X., Warren, K., Kovarcik, D. P., Fellner, T. and Rao, M. S. (2015) ‘cGMP-Manufactured Human Induced Pluripotent Stem Cells Are Available for Pre-clinical and Clinical Applications.’ *Stem cell reports*. 2015/09/24, Elsevier, 5(4) pp. 647–659.

Ban, H., Nishishita, N., Fusaki, N., Tabata, T., Saeki, K., Shikamura, M., Takada, N., Inoue, M., Hasegawa, M., Kawamata, S. and Nishikawa, S.-I. (2011) ‘Efficient generation of transgene-free human induced pluripotent stem cells (iPSCs) by temperature-sensitive Sendai virus vectors.’ *Proceedings of the National Academy of Sciences of the United States of America*. 2011/08/05, National Academy of Sciences, 108(34) pp. 14234–14239.

Bang, J. S., Choi, N. Y., Lee, M., Ko, Kisung, Lee, H. J., Park, Y. S., Jeong, D., Chung, H.-M. and Ko, Kinarm (2018) ‘Optimization of episomal reprogramming for generation of human induced pluripotent stem cells from fibroblasts.’ *Animal Cells and Systems*. Taylor & Francis, 22(2) pp. 132–139.

Bauer, S., Kirschning, C. J., Häcker, H., Redecke, V., Hausmann, S., Akira, S., Wagner, H. and Lipford, G. B. (2001) 'Human TLR9 confers responsiveness to bacterial DNA via species-specific CpG motif recognition.' *Proceedings of the National Academy of Sciences of the United States of America*. 2001/07/24, The National Academy of Sciences, 98(16) pp. 9237–9242.

Baum, C. (2007) 'Insertional mutagenesis in gene therapy and stem cell biology.' *Current opinion in hematology*, 14, August, pp. 337–342.

Beattie, G. M., Lopez, A. D., Bucay, N., Hinton, A., Firpo, M. T., King, C. C. and Hayek, A. (2005) 'Activin A Maintains Pluripotency of Human Embryonic Stem Cells in the Absence of Feeder Layers.' *STEM CELLS*. John Wiley & Sons, Ltd, 23(4) pp. 489–495.

Beers, J., Linask, K. L., Chen, J. A., Siniscalchi, L. I., Lin, Y., Zheng, W., Rao, M. and Chen, G. (2015) 'A cost-effective and efficient reprogramming platform for large-scale production of integration-free human induced pluripotent stem cells in chemically defined culture.' *Scientific reports*. Nature Publishing Group, 5, June, p. 11319.

Biran, A. and Meshorer, E. (2012) 'Concise Review: Chromatin and Genome Organization in Reprogramming.' *STEM CELLS*. John Wiley & Sons, Ltd, 30(9) pp. 1793–1799.

Bissell, M. J., Hall, H. G. and Parry, G. (1982) 'How does the extracellular matrix direct gene expression?' *Journal of Theoretical Biology*, 99(1) pp. 31–68.

Blau, H. M., Chiu, C.-P. and Webster, C. (1983) 'Cytoplasmic activation of human nuclear genes in stable heterocaryons.' *Cell*, 32(4) pp. 1171–1180.

Bolton, W. E. and Barranco, S. C. (1975) 'Characterization of the cell kinetics and growth properties of cystic fibrosis diploid fibroblasts in vitro.' *American journal of human genetics*, 27(3) pp. 394–409.

Boonkaew, B., Tapeng, L., Netsrithong, R., Vatanashevanopakorn, C., Pattanapanyasat, K. and Wattanapanitch, M. (2018) 'Induced pluripotent stem cell line MUSIi006-A derived from hair follicle keratinocytes as a non-invasive somatic cell source.' *Stem Cell Research*. Elsevier B.V., 31, August, pp. 79–82.

Borgohain, M. P., Haridhasapavalan, K. K., Dey, C., Adhikari, P. and Thummer, R. P. (2019) 'An Insight into DNA-free Reprogramming Approaches to Generate Integration-free Induced Pluripotent Stem Cells for Prospective Biomedical Applications.' *Stem Cell Reviews and Reports*. Humana Press Inc. pp. 286–313.

Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., Guenther, M. G., Kumar, R. M., Murray, H. L., Jenner, R. G., Gifford, D. K., Melton, D. A., Jaenisch, R. and Young, R. A. (2005) 'Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells.' *Cell*, 122(6) pp. 947–956.

Brambrink, T., Foreman, R., Welstead, G. G., Lengner, C. J., Wernig, M., Suh, H. and Jaenisch, R. (2008) 'Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells.' *Cell stem cell*, 2(2) pp. 151–159.

Briggs, R. and King, T. J. (1952) 'Transplantation of Living Nuclei From Blastula Cells into Enucleated Frogs' Eggs.' *Proceedings of the National Academy of*

*Sciences of the United States of America*, 38(5) pp. 455–463.

Brown, B. D., Gentner, B., Cantore, A., Colleoni, S., Amendola, M., Zingale, A., Baccarini, A., Lazzari, G., Galli, C. and Naldini, L. (2007) 'Endogenous microRNA can be broadly exploited to regulate transgene expression according to tissue, lineage and differentiation state.' *Nature Biotechnology*, 25(12) pp. 1457–1467.

Buganim, Y., Faddah, D. A., Cheng, A. W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S. L., van Oudenaarden, A. and Jaenisch, R. (2012) 'Single-Cell Expression Analyses during Cellular Reprogramming Reveal an Early Stochastic and a Late Hierarchic Phase.' *Cell*, 150(6) pp. 1209–1222.

Buganim, Y., Faddah, D. A. and Jaenisch, R. (2013) 'Mechanisms and models of somatic cell reprogramming.' *Nature reviews. Genetics*, 14(6) pp. 427–439.

Cantone, I. and Fisher, A. G. (2017) 'Human X chromosome inactivation and reactivation: implications for cell reprogramming and disease.' *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*. The Royal Society, 372(1733) p. 20160358.

Chen, J. (2016) 'The Cell-Cycle Arrest and Apoptotic Functions of p53 in Tumor Initiation and Progression.' *Cold Spring Harbor Perspectives in Medicine*, 6, March, p. a026104.

Chen, Z.-Y., He, C.-Y., Ehrhardt, A. and Kay, M. A. (2003) 'Minicircle DNA vectors devoid of bacterial DNA result in persistent and high-level transgene expression in vivo.' *Molecular Therapy*. Cell Press, 8(3) pp. 495–500.

Chin, M. H., Mason, M. J., Xie, W., Volinia, S., Singer, M., Peterson, C., Ambartsumyan, G., Aimiuwu, O., Richter, L., Zhang, J., Khvorostov, I., Ott, V., Grunstein, M., Lavon, N., Benvenisty, N., Croce, C. M., Clark, A. T., Baxter, T., Pyle, A. D., Teitell, M. A., Pelegrini, M., Plath, K. and Lowry, W. E. (2009) 'Induced Pluripotent Stem Cells and Embryonic Stem Cells Are Distinguished by Gene Expression Signatures.' *Cell Stem Cell*, 5(1) pp. 111–123.

Chung, Y., Klimanskaya, I., Becker, S., Li, T., Maserati, M., Lu, S.-J., Zdravkovic, T., Ilic, D., Genbacev, O., Fisher, S., Krtolica, A. and Lanza, R. (2008) 'Human Embryonic Stem Cell Lines Generated without Embryo Destruction.' *Cell Stem Cell*, 2(2) pp. 113–117.

Churko, J. M., Lee, J., Ameen, M., Gu, M., Venkatasubramanian, M., Diecke, S., Sallam, K., Im, H., Wang, G., Gold, J. D., Salomonis, N., Snyder, M. P. and Wu, J. C. (2017) 'Transcriptomic and epigenomic differences in human induced pluripotent stem cells generated from six reprogramming methods.' *Nature Biomedical Engineering*, 1(10) pp. 826–837.

Cristo, F., Inácio, J. M., Rosas, G., Carreira, I. M., Melo, J. B., de Almeida, L. P., Mendes, P., Martins, D. S., Maio, J., Anjos, R. and Belo, J. A. (2017) 'Generation of human iPSC line from a patient with laterality defects and associated congenital heart anomalies carrying a DAND5 missense alteration.' *Stem Cell Research*. Elsevier B.V., 25, December, pp. 152–156.

Crow, D. (2019) 'Could iPSCs Enable "Off-the-Shelf" Cell Therapy?' *Cell*. Elsevier, 177(7) pp. 1667–1669.

Darquet, A. M., Cameron, B., Wils, P., Scherman, D. and Crouzet, J. (1998) 'A new DNA vehicle for nonviral gene delivery: Supercoiled minicircle.' *Gene therapy*, 4, January, pp. 1341–1349.

David, L. and Polo, J. M. (2014) 'Phases of reprogramming.' *Stem Cell Research*, 12(3) pp. 754–761.

Davis, R. L., Weintraub, H. and Lassar, A. B. (1987) 'Expression of a single transfected cDNA converts fibroblasts to myoblasts.' *Cell*, 51(6) pp. 987–1000.

Davis, R. P., Nemes, C., Varga, E., Freund, C., Kosmidis, G., Gkatzis, K., de Jong, D., Szuhai, K., Dinnyés, A. and Mummery, C. L. (2013) 'Generation of induced pluripotent stem cells from human foetal fibroblasts using the Sleeping Beauty transposon gene delivery system.' *Differentiation*, 86(1) pp. 30–37.

Dempsey, A. and Bowie, A. G. (2015) 'Innate immune recognition of DNA: A recent history.' *Virology*. 2015/03/26, 479–480, May, pp. 146–152.

Deng, Z., Lezina, L., Chen, C.-J., Shtivelband, S., So, W. and Lieberman, P. M. (2002) 'Telomeric proteins regulate episomal maintenance of Epstein-Barr virus origin of plasmid replication.' *Molecular cell*. Elsevier, 9(3) pp. 493–503.

Diecke, S., Lisowski, L., Kooreman, N. G. and Wu, J. C. (2014) 'Second Generation Codon Optimized Minicircle (CoMiC) for Nonviral Reprogramming of Human Adult Fibroblasts BT - Cardiac Tissue Engineering: Methods and Protocols.' *In* Radisic, M. and Black III, L. D. (eds). New York, NY: Springer New York, pp. 1–13.

Downing, T. L., Soto, J., Morez, C., Houssin, T., Fritz, A., Yuan, F., Chu, J., Patel, S., Schaffer, D. V and Li, S. (2013) 'Biophysical regulation of epigenetic state and cell reprogramming.' *Nature Materials*. Nature Publishing Group, 12, October, p. 1154.

Drozd, A. M., Walczak, M. P., Piaskowski, S., Stoczynska-Fidelus, E., Rieske, P. and Grzela, D. P. (2015) 'Generation of human iPSCs from cells of fibroblastic and epithelial origin by means of the oriP/EBNA-1 episomal reprogramming system.' *Stem cell research & therapy*. BioMed Central, 6(1) p. 122.

Eggenberger, J., Blanco-Melo, D., Panis, M., Brennand, K. J. and tenOever, B. R. (2019) 'Type I interferon response impairs differentiation potential of pluripotent stem cells.' *Proceedings of the National Academy of Sciences*, 116(4) pp. 1384 LP – 1393.

Eguizabal, C., Aran, B., Chuva de Sousa Lopes, S. M., Geens, M., Heindryckx, B., Panula, S., Popovic, M., Vassena, R. and Veiga, A. (2019) 'Two decades of embryonic stem cells: a historical overview.' *Human reproduction open*. Oxford University Press, 2019(1) pp. hoy024–hoy024.

Frappier, L. (2012) 'The Epstein-Barr Virus EBNA1 Protein.' *Scientifica*. 2012/12/19, Hindawi Publishing Corporation, 2012 p. 438204.

Fusaki, N., Ban, H., Nishiyama, A., Saeki, K. and Hasegawa, M. (2009) 'Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome.' *Proceedings of the Japan Academy. Series B, Physical and biological sciences*. The Japan Academy, 85(8) pp. 348–362.

Gaczynska, M., Rock, K. L. and Goldberg, A. L. (1993) 'γ-Interferon and expression of MHC genes regulate peptide hydrolysis by proteasomes.' *Nature*, 365(6443) pp. 264–267.

*Glossary* | *stemcells.nih.gov* (n.d.). [Online] [Accessed on 18th September 2019] https://stemcells.nih.gov/glossary.htm#embryonicsc.

González, F., Boué, S. and Belmonte, J. C. I. (2011) 'Methods for making induced pluripotent stem cells: reprogramming à la carte.' *Nature Reviews Genetics*, 12(4) pp. 231–242.

Gore, A., Li, Z., Fung, H.-L., Young, J. E., Agarwal, S., Antosiewicz-Bourget, J., Canto, I., Giorgetti, A., Israel, M. A., Kiskinis, E., Lee, J.-H., Loh, Y.-H., Manos, P. D., Montserrat, N., Panopoulos, A. D., Ruiz, S., Wilbert, M. L., Yu, J., Kirkness, E. F., Belmonte, J. C. I., Rossi, D. J., Thomson, J. A., Eggan, K., Daley, G. Q., Goldstein, L. S. B. and Zhang, K. (2011) 'Somatic coding mutations in human induced pluripotent stem cells.' *Nature*, 471(7336) pp. 63–67.

Gruhne, B., Sompallae, R., Marescotti, D., Kamranvar, S. A., Gastaldello, S. and Masucci, M. G. (2009) 'The Epstein–Barr virus nuclear antigen-1 promotes genomic instability via induction of reactive oxygen species.' *Proceedings of the National Academy of Sciences*, 106(7) pp. 2313 LP – 2318.

Gurdon, J. B. (1962) 'The Developmental Capacity of Nuclei taken from Intestinal Epithelium Cells of Feeding Tadpoles.' *Journal of Embryology and Experimental Morphology*, 10(4) pp. 622 LP – 640.

Hacein-Bey-Abina, S., Kalle, C., Schmidt, M., Deist, F., Wulffraat, N., McIntyre, E., Radford, I., Villeval, J.-L., Fraser, C., Cavazzana-Calvo, M. and Fischer, A. (2003) 'A Serious Adverse Event after Successful Gene Therapy for X-Linked Severe Combined Immunodeficiency.' *The New England journal of medicine*, 348, February, pp. 255–256.

Haridhasapavalan, K. K., Borgohain, M. P., Dey, C., Saha, B., Narayan, G., Kumar, S. and Thummer, R. P. (2019) 'An insight into non-integrative gene delivery approaches to generate transgene-free induced pluripotent stem cells.' *Gene*, 686 pp. 146–159.

Hawkins, K. E., Joy, S., Delhove, J. M. K. M., Kotiadis, V. N., Fernandez, E., Fitzpatrick, L. M., Whiteford, J. R., King, P. J., Bolanos, J. P., Duchen, M. R., Waddington, S. N. and McKay, T. R. (2016) 'NRF2 Orchestrates the Metabolic Shift during Induced Pluripotent Stem Cell Reprogramming.' *Cell reports*. 2016/02/18, Cell Press, 14(8) pp. 1883–1891.

Hemmi, H., Takeuchi, O., Kawai, T., Kaisho, T., Sato, S., Sanjo, H., Matsumoto, M., Hoshino, K., Wagner, H., Takeda, K. and Akira, S. (2000) 'A Toll-like receptor recognizes bacterial DNA.' *Nature*, 408(6813) pp. 740–745.

Higuchi, A., Ling, Q.-D., Ko, Y.-A., Chang, Y. and Umezawa, A. (2011) 'Biomaterials for the Feeder-Free Culture of Human Embryonic Stem Cells and Induced Pluripotent Stem Cells.' *Chemical Reviews*. American Chemical Society, 111(5) pp. 3021–3035.

Hodin, T. L., Najrana, T. and Yates, J. L. (2013) 'Efficient Replication of Epstein-Barr

Virus-Derived Plasmids Requires Tethering by EBNA1 to Host Chromosomes.' *Journal of Virology*, 87(23) pp. 13020 LP – 13028.

Hong, H., Takahashi, K., Ichisaka, T., Aoi, T., Kanagawa, O., Nakagawa, M., Okita, K. and Yamanaka, S. (2009) 'Suppression of induced pluripotent stem cell generation by the p53–p21 pathway.' *Nature*. Macmillan Publishers Limited. All rights reserved, 460, August, p. 1132.

Hong, M., Murai, Y., Kutsuna, T., Takahashi, H., Nomoto, K., Cheng, C.-M., Ishizawa, S., Zhao, Q.-L., Ogawa, R., Harmon, B., Tsuneyama, K. and Takano, Y. (2006) 'Suppression of Epstein-Barr nuclear antigen 1 (EBNA1) by RNA interference inhibits proliferation of EBV-positive Burkitt's lymphoma cells.' *Journal of cancer research and clinical oncology*, 132, February, pp. 1–8.

Horikawa, I., Park, K., Isogaya, K., Hiyoshi, Y., Li, H., Anami, K., Robles, A. I., Mondal, A. M., Fujita, K., Serrano, M. and Harris, C. C. (2017) 'Δ133p53 represses p53-inducible senescence genes and enhances the generation of human induced pluripotent stem cells.' *Cell Death & Differentiation*, 24(6) pp. 1017–1028.

Hu, K. (2014a) 'All roads lead to induced pluripotent stem cells: the technologies of iPSC generation.' *Stem cells and development*. 2014/03/21, Mary Ann Liebert, Inc., 23(12) pp. 1285–1300.

Hu, K. (2014b) 'Vectorology and Factor Delivery in Induced Pluripotent Stem Cell Reprogramming.' *Stem Cells and Development*. Mary Ann Liebert, Inc., publishers, 23(12) pp. 1301–1315.

Igawa, K., Kokubu, C., Yusa, K., Horie, K., Yoshimura, Y., Yamauchi, K., Suemori, H., Yokozeki, H., Toyoda, M., Kiyokawa, N., Okita, H., Miyagawa, Y., Akutsu, H., Umezawa, A., Katayama, I. and Takeda, J. (2014) 'Removal of Reprogramming Transgenes Improves the Tissue Reconstitution Potential of Keratinocytes Generated From Human Induced Pluripotent Stem Cells.' *STEM CELLS Translational Medicine*. John Wiley & Sons, Ltd, 3(9) pp. 992–1001.

Jemielity, J., Fowler, T., Zuberek, J., Stepinski, J., Lewdorowicz, M., Niedzwiecka, A., Stolarski, R., Darzynkiewicz, E. and Rhoads, R. E. (2003) 'Novel "anti-reverse" cap analogs with superior translational properties.' *RNA (New York, N.Y.)*. Copyright 2003 by RNA Society, 9(9) pp. 1108–1122.

Jia, F., Wilson, K. D., Sun, N., Gupta, D. M., Huang, M., Li, Z., Panetta, N. J., Chen, Z. Y., Robbins, R. C., Kay, M. A., Longaker, M. T. and Wu, J. C. (2010) 'A nonviral minicircle vector for deriving human iPS cells.' *Nature methods*. 2010/02/07, 7(3) pp. 197–199.

Kaji, K., Norrby, K., Paca, A., Mileikovsky, M., Mohseni, P. and Woltjen, K. (2009) 'Virus-free induction of pluripotency and subsequent excision of reprogramming factors.' *Nature*, 458(7239) pp. 771–775.

Kapinas, K., Grandy, R., Ghule, P., Medina, R., Becker, K., Pardee, A., Zaidi, S. K., Lian, J., Stein, J., van Wijnen, A. and Stein, G. (2013) 'The abbreviated pluripotent cell cycle.' *Journal of cellular physiology*, 228(1) pp. 9–20.

Karbowniczek, K., Rothwell, P., Extance, J., Milsom, S., Lukashchuk, V., Bowes, K., Smith, D. and Caproni, L. (2017) 'Doggybone™ DNA: an advanced platform for AAV

production.' *Cell and Gene Therapy Insights*, 3, November, pp. 731–738.

Karikó, K., Kuo, A. and Barnathan, E. S. (1999) 'Overexpression of urokinase receptor in mammalian cells following administration of the in vitro transcribed encoding mRNA.' *Gene Therapy*, 6(6) pp. 1092–1100.

Karikó, K., Muramatsu, H., Welsh, F. A., Ludwig, J., Kato, H., Akira, S. and Weissman, D. (2008) 'Incorporation of pseudouridine into mRNA yields superior nonimmunogenic vector with increased translational capacity and biological stability.' *Molecular therapy : the journal of the American Society of Gene Therapy*. 2008/09/16, 16(11) pp. 1833–1840.

Kastin, A. (2013) *Handbook of Biologically Active Peptides*.

Kawasaki, T. and Kawai, T. (2014) 'Toll-Like Receptor Signaling Pathways .' *Frontiers in Immunology* p. 461.

Kennedy, G., Komano, J. and Sugden, B. (2003) 'Epstein-Barr virus provides a survival factor to Burkitt's lymphomas.' *Proceedings of the National Academy of Sciences*, 100(24) pp. 14269 LP – 14274.

Kikuchi, T., Morizane, A., Doi, D., Magotani, H., Onoe, H., Hayashi, T., Mizuma, H., Takara, S., Takahashi, R., Inoue, H., Morita, S., Yamamoto, M., Okita, K., Nakagawa, M., Parmar, M. and Takahashi, J. (2017) 'Human iPS cell-derived dopaminergic neurons function in a primate Parkinson's disease model.' *Nature*. Macmillan Publishers Limited, part of Springer Nature. All rights reserved., 548, August, p. 592.

Kim, D., Kim, C.-H., Moon, J.-I., Chung, Y.-G., Chang, M.-Y., Han, B.-S., Ko, S., Yang, E., Cha, K. Y., Lanza, R. and Kim, K.-S. (2009) 'Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins.' *Cell stem cell*. 2009/05/28, 4(6) pp. 472–476.

Kim, K.-Y., Hysolli, E., Tanaka, Y., Wang, B., Jung, Y.-W., Pan, X., Weissman, S. M. and Park, I.-H. (2014) 'X Chromosome of female cells shows dynamic changes in status during human somatic cell reprogramming.' *Stem cell reports*. Elsevier, 2(6) pp. 896–909.

Lang, R., Liu, G., Shi, Y., Bharadwaj, S., Leng, X., Zhou, X., Liu, H., Atala, A. and Zhang, Y. (2013) 'Self-renewal and differentiation capacity of urine-derived stem cells after urine preservation for 24 hours.' *PloS one*. 2013/01/18, Public Library of Science, 8(1) pp. e53980–e53980.

Lee, J. H., Yu, H.-S., Lee, G.-S., Ji, A., Hyun, J. K. and Kim, H.-W. (2011) 'Collagen gel three-dimensional matrices combined with adhesive proteins stimulate neuronal differentiation of mesenchymal stem cells.' *Journal of the Royal Society, Interface*. 2011/01/19, The Royal Society, 8(60) pp. 998–1010.

Lee, J., Sayed, N., Hunter, A., Au, K. F., Wong, W. H., Mocarski, E. S., Pera, R. R., Yakubov, E. and Cooke, J. P. (2012) 'Activation of innate immunity is required for efficient nuclear reprogramming.' *Cell*, 151(3) pp. 547–558.

Lee, K.-I., Kim, H.-T. and Hwang, D.-Y. (2014) 'Footprint- and xeno-free human iPSCs derived from urine cells using extracellular matrix-based culture conditions.' *Biomaterials*. Elsevier, 35(29) pp. 8330–8338.

Lindner, S. E. and Sugden, B. (2007) ‘The plasmid replicon of Epstein-Barr virus: mechanistic insights into efficient, licensed, extrachromosomal replication in human cells.’ *Plasmid*. 2007/03/09, 58(1) pp. 1–12.

Llames, S., García-Pérez, E., Meana, Á., Larcher, F. and del Río, M. (2015) ‘Feeder Layer Cell Actions and Applications.’ *Tissue engineering. Part B, Reviews*. 2015/03/24, Mary Ann Liebert, Inc., 21(4) pp. 345–353.

Maherali, N. and Hochedlinger, K. (2008) ‘Guidelines and techniques for the generation of induced pluripotent stem cells.’ *Cell stem cell*. Elsevier, 3(6) pp. 595–605.

Mahla, R. S. (2016) ‘Stem Cells Applications in Regenerative Medicine and Disease Therapeutics.’ *International journal of cell biology*. 2016/07/19, Hindawi Publishing Corporation, 2016 p. 6940283.

Mandai, M., Watanabe, A., Kurimoto, Y., Hirami, Y., Morinaga, C., Daimon, T., Fujihara, M., Akimaru, H., Sakai, N., Shibata, Y., Terada, M., Nomiya, Y., Tanishima, S., Nakamura, M., Kamao, H., Sugita, S., Onishi, A., Ito, T., Fujita, K. and Takahashi, M. (2017) ‘Autologous Induced Stem-Cell–Derived Retinal Cells for Macular Degeneration.’ *New England Journal of Medicine*, 376, March, pp. 1038–1046.

Marión, R. M., Strati, K., Li, H., Murga, M., Blanco, R., Ortega, S., Fernandez-Capetillo, O., Serrano, M. and Blasco, M. A. (2009) ‘A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity.’ *Nature*. 2009/08/09, 460(7259) pp. 1149–1153.

Martin, G. R. (1981) ‘Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells.’ *Proceedings of the National Academy of Sciences of the United States of America*, 78(12) pp. 7634–7638.

Maucksch, C., Bohla, A., Hoffmann, F., Schleef, M., Aneja, M. K., Elfinger, M., Hartl, D. and Rudolph, C. (2009) ‘Transgene expression of transfected supercoiled plasmid DNA concatemers in mammalian cells.’ *The Journal of Gene Medicine*. John Wiley & Sons, Ltd, 11(5) pp. 444–453.

McLenachan, S., Sarsero, J. P. and Ioannou, P. A. (2007) ‘Flow-cytometric analysis of mouse embryonic stem cell lipofection using small and large DNA constructs.’ *Genomics*. Academic Press, 89(6) pp. 708–720.

Merkle, F. T., Ghosh, S., Kamitaki, N., Mitchell, J., Avior, Y., Mello, C., Kashin, S., Mekhoubad, S., Ilic, D., Charlton, M., Saphier, G., Handsaker, R. E., Genovese, G., Bar, S., Benvenisty, N., McCarroll, S. A. and Eggan, K. (2017) ‘Human pluripotent stem cells recurrently acquire and expand dominant negative P53 mutations.’ *Nature*, 545(7653) pp. 229–233.

Miyoshi, N., Ishii, H., Nagano, H., Haraguchi, N., Dewi, D. L., Kano, Y., Nishikawa, S., Tanemura, M., Mimori, K., Tanaka, F., Saito, T., Nishimura, J., Takemasa, I., Mizushima, T., Ikeda, M., Yamamoto, H., Sekimoto, M., Doki, Y. and Mori, M. (2011) ‘Reprogramming of Mouse and Human Cells to Pluripotency Using Mature MicroRNAs.’ *Cell Stem Cell*. Cell Press, 8(6) pp. 633–638.

Mockey, M., Gonçalves, C., Dupuy, F. P., Lemoine, F. M., Pichon, C. and Midoux, P.

(2006) ‘mRNA transfection of dendritic cells: Synergistic effect of ARCA mRNA capping with Poly(A) chains in cis and in trans for a high protein expression level.’ *Biochemical and Biophysical Research Communications*. Academic Press, 340(4) pp. 1062–1068.

Monti, M., Perotti, C., del fante, C. and Redi, C. (2012) ‘Stem cells: Sources and therapies.’ *Biological research*, 45, January, pp. 207–214.

Mora, C., Serzanti, M., Consiglio, A., Memo, M. and Dell’Era, P. (2017) ‘Clinical potentials of human pluripotent stem cells.’ *Cell Biology and Toxicology*, 33, February.

Moriyama, K., Yoshizawa-Sugata, N., Obuse, C., Tsurimoto, T. and Masai, H. (2012) ‘Epstein-Barr nuclear antigen 1 (EBNA1)-dependent recruitment of origin recognition complex (Orc) on oriP of Epstein-Barr virus with purified proteins: stimulation by Cdc6 through its direct interaction with EBNA1.’ *The Journal of biological chemistry*. 2012/05/14, American Society for Biochemistry and Molecular Biology, 287(28) pp. 23977–23994.

Muenthaisong, S., Ujhelly, O., Polgar, Z., Varga, E., Ivics, Z., Pirity, M. K. and Dinnyes, A. (2012) ‘Generation of mouse induced pluripotent stem cells from different genetic backgrounds using Sleeping beauty transposon mediated gene transfer.’ *Experimental Cell Research*. Academic Press, 318(19) pp. 2482–2489.

Nagoshi, N., Tsuji, O., Nakamura, M. and Okano, H. (2019) ‘Cell therapy for spinal cord injury using induced pluripotent stem cells.’ *Regenerative Therapy*. Elsevier, 11, December, pp. 75–80.

Nakanishi, M. and Otsu, M. (2012) ‘Development of Sendai virus vectors and their potential applications in gene therapy and regenerative medicine.’ *Current gene therapy*. Bentham Science Publishers, 12(5) pp. 410–416.

Narsinh, K. H., Jia, F., Robbins, R. C., Kay, M. A., Longaker, M. T. and Wu, J. C. (2011) ‘Generation of adult human induced pluripotent stem cells using nonviral minicircle DNA vectors.’ *Nature Protocols*, 6(1) pp. 78–88.

Nesbit, C. E., Tersak, J. M. and Prochownik, E. V (1999) ‘MYC oncogenes and human neoplastic disease.’ *Oncogene*, 18(19) pp. 3004–3016.

Niakan, K. K., Han, J., Pedersen, R. A., Simon, C. and Pera, R. A. R. (2012) ‘Human pre-implantation embryo development.’ *Development (Cambridge, England)*. Company of Biologists, 139(5) pp. 829–841.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Schöler, H. and Smith, A. (1998) ‘Formation of Pluripotent Stem Cells in the Mammalian Embryo Depends on the POU Transcription Factor Oct4.’ *Cell*, 95(3) pp. 379–391.

Nishimura, K., Ohtaka, M., Takada, H., Kurisaki, A., Tran, N. V. K., Tran, Y. T. H., Hisatake, K., Sano, M. and Nakanishi, M. (2017) ‘Simple and effective generation of transgene-free induced pluripotent stem cells using an auto-erasable Sendai virus vector responding to microRNA-302.’ *Stem Cell Research*, 23 pp. 13–19.

Nusse, R. (2008) ‘Wnt signaling and stem cell control.’ *Cell Research*, 18(5) pp. 523–527.

O'Neill, L. A. J., Golenbock, D. and Bowie, A. G. (2013) 'The history of Toll-like receptors — redefining innate immunity.' *Nature Reviews Immunology*, 13(6) pp. 453–460.

Okita, K., Ichisaka, T. and Yamanaka, S. (2007) 'Generation of germline-competent induced pluripotent stem cells.' *Nature*, 448(7151) pp. 313–317.

Okita, K., Matsumura, Y., Sato, Y., Okada, A., Morizane, A., Okamoto, S., Hong, H., Nakagawa, M., Tanabe, K., Tezuka, K., Shibata, T., Kunisada, T., Takahashi, M., Takahashi, J., Saji, H. and Yamanaka, S. (2011) 'A more efficient method to generate integration-free human iPS cells.' *Nature Methods*, 8(5) pp. 409–412.

Okita, K., Nakagawa, M., Hyenjong, H., Ichisaka, T. and Yamanaka, S. (2008) 'Generation of Mouse Induced Pluripotent Stem Cells Without Viral Vectors.' *Science*, 322(5903) pp. 949 LP – 953.

Okita, K., Yamakawa, T., Matsumura, Y., Sato, Y., Amano, N., Watanabe, A., Goshima, N. and Yamanaka, S. (2013) 'An Efficient Nonviral Method to Generate Integration-Free Human-Induced Pluripotent Stem Cells from Cord Blood and Peripheral Blood Cells.' *STEM CELLS*. John Wiley & Sons, Ltd, 31(3) pp. 458–466.

Osborn, M. J., Panoskaltsis-Mortari, A., McElmurry, R. T., Bell, S. K., Vignali, D. A. A., Ryan, M. D., Wilber, A. C., McIvor, R. S., Tolar, J. and Blazar, B. R. (2005) 'A picornaviral 2A-like sequence-based tricistronic vector allowing for high-level therapeutic gene expression coupled to a dual-reporter system.' *Molecular therapy : the journal of the American Society of Gene Therapy*. Elsevier, 12(3) pp. 569–74.

Pannone, G., Zamparese, R., Pace, M., Pedicillo, M. C., Cagiano, S., Somma, P., Errico, M. E., Donofrio, V., Franco, R., De Chiara, A., Aquino, G., Bucci, P., Bucci, E., Santoro, A. and Bufo, P. (2014) 'The role of EBV in the pathogenesis of Burkitt's Lymphoma: an Italian hospital based survey.' *Infectious agents and cancer*. BioMed Central, 9(1) p. 34.

Qin, L., Ding, Y., Pahud, D. R., Chang, E., Imperiale, M. J. and Bromberg, J. S. (1997) 'Promoter Attenuation in Gene Therapy: Interferon-γ and Tumor Necrosis Factor-α Inhibit Transgene Expression.' *Human Gene Therapy*. Mary Ann Liebert, Inc., publishers, 8(17) pp. 2019–2029.

Ramalho-Santos, M. and Willenbring, H. (2007) 'On the Origin of the Term “Stem Cell”' *Cell Stem Cell*. Elsevier, 1(1) pp. 35–38.

Rao, M. S. and Malik, N. (2012) 'Assessing iPSC reprogramming methods for their suitability in translational medicine.' *Journal of cellular biochemistry*, 113(10) pp. 3061–3068.

Rasmussen, M. A., Holst, B., Tümer, Z., Johnsen, M. G., Zhou, S., Stummann, T. C., Hyttel, P. and Clausen, C. (2014) 'Transient p53 Suppression Increases Reprogramming of Human Fibroblasts without Affecting Apoptosis and DNA Damage.' *Stem Cell Reports*. Elsevier, 3(3) pp. 404–413.

Redon, C. E., Dickey, J. S., Bonner, W. M. and Sedelnikova, O. A. (2009) 'γ-H2AX as a biomarker of DNA damage induced by ionizing radiation in human peripheral blood lymphocytes and artificial skin.' *Advances in space research : the official journal of the Committee on Space Research (COSPAR)*, 43(8) pp. 1171–1178.

Richards, M., Fong, C.-Y., Chan, W.-K., Wong, P.-C. and Bongso, A. (2002) 'Human feeders support prolonged undifferentiated growth of human inner cell masses and embryonic stem cells.' *Nature Biotechnology*, 20(9) pp. 933–936.

Romorini, L., Garate, X., Neiman, G., Luzzani, C., Furmento, V. A., Guberman, A. S., Sevlever, G. E., Scassa, M. E. and Miriuka, S. G. (2016) 'AKT/GSK3β signaling pathway is critically involved in human pluripotent stem cell survival.' *Scientific reports*. Nature Publishing Group, 6, October, p. 35660.

Saha, S., Ji, L., de Pablo, J. J. and Palecek, S. P. (2008) 'TGFbeta/Activin/Nodal pathway in inhibition of human embryonic stem cell differentiation by mechanical strain.' *Biophysical journal*. 2008/01/30, The Biophysical Society, 94(10) pp. 4123–4133.

Samavarchi-Tehrani, P., Golipour, A., David, L., Sung, H., Beyer, T. A., Datti, A., Woltjen, K., Nagy, A. and Wrana, J. L. (2010) 'Functional Genomics Reveals a BMP-Driven Mesenchymal-to-Epithelial Transition in the Initiation of Somatic Cell Reprogramming.' *Cell Stem Cell*, 7(1) pp. 64–77.

Schepers, A., Ritzi, M., Bousset, K., Kremmer, E., Yates, J. L., Harwood, J., Diffley, J. F. and Hammerschmidt, W. (2001) 'Human origin recognition complex binds to the region of the latent origin of DNA replication of Epstein-Barr virus.' *The EMBO journal*. Oxford University Press, 20(16) pp. 4588–4602.

Schlaeger, T. M., Daheron, L., Brickler, T. R., Entwisle, S., Chan, K., Cianci, A., DeVine, A., Ettenger, A., Fitzgerald, K., Godfrey, M., Gupta, D., McPherson, J., Malwadkar, P., Gupta, M., Bell, B., Doi, A., Jung, N., Li, X., Lynes, M. S., Brookes, E., Cherry, A. B. C., Demirbas, D., Tsankov, A. M., Zon, L. I., Rubin, L. L., Feinberg, A. P., Meissner, A., Cowan, C. A. and Daley, G. Q. (2015) 'A comparison of non-integrating reprogramming methods.' *Nature biotechnology*. 2014/12/01, 33(1) pp. 58–63.

Sears, J., Ujihara, M., Wong, S., Ott, C., Middeldorp, J. and Aiyar, A. (2004) 'The Amino Terminus of Epstein-Barr Virus (EBV) Nuclear Antigen 1 Contains AT Hooks That Facilitate the Replication and Partitioning of Latent EBV Genomes by Tethering Them to Cellular Chromosomes.' *Journal of Virology*, 78(21) pp. 11487 LP – 11505.

Seki, T., Yuasa, S., Oda, M., Egashira, T., Yae, K., Kusumoto, D., Nakata, H., Tohyama, S., Hashimoto, H., Kodaira, M., Okada, Y., Seimiya, H., Fusaki, N., Hasegawa, M. and Fukuda, K. (2010) 'Generation of Induced Pluripotent Stem Cells from Human Terminally Differentiated Circulating T Cells.' *Cell Stem Cell*. Elsevier, 7(1) pp. 11–14.

Si-Tayeb, K., Noto, F. K., Sepac, A., Sedlic, F., Bosnjak, Z. J., Lough, J. W. and Duncan, S. A. (2010) 'Generation of human induced pluripotent stem cells by simple transient transfection of plasmid DNA encoding reprogramming factors.' *BMC Developmental Biology*, 10(1) p. 81.

Smith, Z. D., Nachman, I., Regev, A. and Meissner, A. (2010) 'Dynamic single-cell imaging of direct reprogramming reveals an early specifying event.' *Nature Biotechnology*, 28(5) pp. 521–526.

Stadtfeld, M. and Hochedlinger, K. (2010) 'Induced pluripotency: history, mechanisms, and applications.' *Genes & development*. Cold Spring Harbor

Laboratory Press, 24(20) pp. 2239–2263.

Steimle, V., Siegrist, C. A., Mottet, A., Lisowska-Grospierre, B. and Mach, B. (1994) 'Regulation of MHC class II expression by interferon-gamma mediated by the transactivator gene CIITA.' *Science*, 265(5168) pp. 106 LP – 109.

Sullivan, S., Stacey, G. N., Akazawa, C., Aoyama, N., Baptista, R., Bedford, P., Bennaceur Griscelli, A., Chandra, A., Elwood, N., Girard, M., Kawamata, S., Hanatani, T., Latsis, T., Lin, S., Ludwig, T. E., Malygina, T., Mack, A., Mountford, J. C., Noggle, S., Pereira, L. V, Price, J., Sheldon, M., Srivastava, A., Stachelscheid, H., Velayudhan, S. R., Ward, N. J., Turner, M. L., Barry, J. and Song, J. (2018) 'Quality control guidelines for clinical-grade human induced pluripotent stem cell lines.' *Regenerative Medicine*. Future Medicine, 13(7) pp. 859–866.

Tada, M., Takahama, Y., Abe, K., Nakatsuji, N. and Tada, T. (2001) 'Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells.' *Current Biology*, 11(19) pp. 1553–1558.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K. and Yamanaka, S. (2007) 'Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors.' *Cell*, 131(5) pp. 861–872.

Takahashi, K. and Yamanaka, S. (2016) 'A decade of transcription factor-mediated reprogramming to pluripotency.' *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 17, February, p. 183.

Tanaka, Y., Hysolli, E., Su, J., Xiang, Y., Kim, K.-Y., Zhong, M., Li, Y., Heydari, K., Euskirchen, G., Snyder, M. P., Pan, X., Weissman, S. M. and Park, I.-H. (2015) 'Transcriptome Signature and Regulation in Human Somatic Cell Reprogramming.' *Stem Cell Reports*, 4(6) pp. 1125–1139.

*The Germ-plasm: A Theory of Heredity : August Weismann : Free Download, Borrow, and Streaming : Internet Archive* (n.d.). [Online] [Accessed on 25th September 2019] https://archive.org/details/germplasmatheor02weisgoog/page/n6.

Thelemann, C., Eren, R. O., Coutaz, M., Brasseit, J., Bouzourene, H., Rosa, M., Duval, A., Lavanchy, C., Mack, V., Mueller, C., Reith, W. and Acha-Orbea, H. (2014) 'Interferon-γ Induces Expression of MHC Class II on Intestinal Epithelial Cells and Protects Mice from Colitis.' *PLOS ONE*. Public Library of Science, 9(1) p. e86844.

Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S. and Jones, J. M. (1998) 'Embryonic Stem Cell Lines Derived from Human Blastocysts.' *Science*, 282(5391) pp. 1145 LP – 1147.

Tokuzawa, Y., Kaiho, E., Maruyama, M., Takahashi, K., Mitsui, K., Maeda, M., Niwa, H. and Yamanaka, S. (2003) 'Fbx15 is a novel target of Oct3/4 but is dispensable for embryonic stem cell self-renewal and mouse development.' *Molecular and cellular biology*. American Society for Microbiology, 23(8) pp. 2699–2708.

Uryga, A., Gray, K. and Bennett, M. (2015) 'DNA Damage and Repair in Vascular Disease.' *Annual review of physiology*, 78, October.

Valentine, R., Dawson, C. W., Hu, C., Shah, K. M., Owen, T. J., Date, K. L., Maia, S. P., Shao, J., Arrand, J. R., Young, L. S. and O'Neil, J. D. (2010) 'Epstein-Barr virus-

encoded EBNA1 inhibits the canonical NF-kappaB pathway in carcinoma cells by inhibiting IKK phosphorylation.' *Molecular cancer*. BioMed Central, 9, January, p. 1.

Vanneaux, V. (2019) 'Induced Pluripotent Stem Cells for Clinical Use.' *In Update on Induced Pluripotent Stem Cells [Working Title]*. IntechOpen.

Vazin, T. and Freed, W. J. (2010) 'Human embryonic stem cells: derivation, culture, and differentiation: a review.' *Restorative neurology and neuroscience*, 28(4) pp. 589–603.

Waddington, C. H. (1957) *The strategy of the genes: a discussion of some aspects of theoretical biology*. Allen & Unwin.

Wang, Lihui, Wang, Linli, Huang, W., Su, H., Xue, Y., Su, Z., Liao, B., Wang, H., Bao, X., Qin, D., He, J., Wu, W., So, K. F., Pan, G. and Pei, D. (2013) 'Generation of integration-free neural progenitor cells from cells in human urine.' *Nature Methods*, 10(1) pp. 84–89.

Wang, W., Jia, Y., Li, Y., Jing, C., Guo, X., Shang, X., Zhao, C. and Wang, T. (2017) 'RETRACTED ARTICLE: Impact of different promoters, promoter mutation, and an enhancer on recombinant protein expression in CHO cells.' *Scientific Reports*, 7(1) p. 10416.

Wang, W., Lin, C., Lu, D., Ning, Z., Cox, T., Melvin, D., Wang, X., Bradley, A. and Liu, P. (2008) 'Chromosomal transposition of <em>PiggyBac</em> in mouse embryonic stem cells.' *Proceedings of the National Academy of Sciences*, 105(27) pp. 9290 LP – 9295.

Wang, Y., Wang, F., Wang, R., Zhao, P. and Xia, Q. (2015) '2A self-cleaving peptide-based multi-gene expression system in the silkworm Bombyx mori.' *Scientific reports*. Nature Publishing Group, 5, November, p. 16273.

Warren, L. and Lin, C. (2019) 'mRNA-Based Genetic Reprogramming.' *Molecular therapy : the journal of the American Society of Gene Therapy*. Elsevier, 27(4) pp. 729–734.

Warren, L., Manos, P. D., Ahfeldt, T., Loh, Y.-H., Li, H., Lau, F., Ebina, W., Mandal, P. K., Smith, Z. D., Meissner, A., Daley, G. Q., Brack, A. S., Collins, J. J., Cowan, C., Schlaeger, T. M. and Rossi, D. J. (2010) 'Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA.' *Cell stem cell*. 2010/09/30, 7(5) pp. 618–630.

Weltner, J., Balboa, D., Katayama, S., Bespalov, M., Krjutškov, K., Jouhilahti, E.-M., Trokovic, R., Kere, J. and Otonkoski, T. (2018) 'Human pluripotent reprogramming with CRISPR activators.' *Nature communications*. Nature Publishing Group UK, 9(1) p. 2643.

Wilmut, I., Schnieke, A. E., McWhir, J., Kind, A. J. and Campbell, K. H. S. (1997) 'Viable offspring derived from fetal and adult mammalian cells.' *Nature*, 385(6619) pp. 810–813.

Woltjen, K., Michael, I. P., Mohseni, P., Desai, R., Mileikovsky, M., Hämäläinen, R., Cowling, R., Wang, W., Liu, P., Gertsenstein, M., Kaji, K., Sung, H.-K. and Nagy, A. (2009) 'piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells.' *Nature*, 458(7239) pp. 766–770.

Wood, V. H. J., O'Neil, J. D., Wei, W., Stewart, S. E., Dawson, C. W. and Young, L. S. (2007) 'Epstein–Barr virus-encoded EBNA1 regulates cellular gene transcription and modulates the STAT1 and TGFβ signaling pathways.' *Oncogene*, 26(28) pp. 4135–4147.

Yakubov, E., Rechavi, G., Rozenblatt, S. and Givol, D. (2010) 'Reprogramming of human fibroblasts to pluripotent stem cells using mRNA of four transcription factors.' *Biochemical and Biophysical Research Communications*. Academic Press, 394(1) pp. 189–193.

Yang, W. (2014) 'iPSC reprogramming from human peripheral blood using Sendai Virus mediated gene transfer.' *StemBook*. Stem Book.

Yates, J. L., Camiolo, S. M. and Bashaw, J. M. (2000) 'The minimal replicator of Epstein-Barr virus oriP.' *Journal of virology*. American Society for Microbiology, 74(10) pp. 4512–4522.

Yew, N. S., Zhao, H., Wu, I.-H., Song, A., Tousignant, J. D., Przybylska, M. and Cheng, S. H. (2000) 'Reduced Inflammatory Response to Plasmid DNA Vectors by Elimination and Inhibition of Immunostimulatory CpG Motifs.' *Molecular Therapy*. Cell Press, 1(3) pp. 255–262.

Yi, L., Lu, C., Hu, W., Sun, Y. and Levine, A. J. (2012) 'Multiple Roles of p53-Related Pathways in Somatic Cell Reprogramming and Stem Cell Differentiation.' *Cancer Research*, 72(21) pp. 5635 LP – 5645.

Yu, J., Hu, K., Smuga-Otto, K., Tian, S., Stewart, R., Slukvin, I. I. and Thomson, J. A. (2009) 'Human induced pluripotent stem cells free of vector and transgene sequences.' *Science (New York, N.Y.)*. 2009/03/26, 324(5928) pp. 797–801.

Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., Nie, J., Jonsdottir, G. A., Ruotti, V., Stewart, R., Slukvin, I. I. and Thomson, J. A. (2007) 'Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells.' *Science*, 318(5858) pp. 1917 LP – 1920.

Zakrzewski, W., Dobrzyński, M., Szymonowicz, M. and Rybak, Z. (2019) 'Stem cells: past, present, and future.' *Stem Cell Research & Therapy*, 10(1) p. 68.

Zhang, M., Yang, C., Liu, H. and Sun, Y. (2013) 'Induced Pluripotent Stem Cells Are Sensitive to DNA Damage.' *Genomics, Proteomics & Bioinformatics*, 11(5) pp. 320–326.

Zhao, T. and Xu, Y. (2010) 'p53 and stem cells: new developments and new concerns.' *Trends in Cell Biology*, 20(3) pp. 170–175.

Zhou, H., Wu, S., Joo, J. Y., Zhu, S., Han, D. W., Lin, T., Trauger, S., Bien, G., Yao, S., Zhu, Y., Siuzdak, G., Schöler, H. R., Duan, L. and Ding, S. (2009) 'Generation of induced pluripotent stem cells using recombinant proteins.' *Cell stem cell*. Elsevier, 4(5) pp. 381–4.

Zhou, T., Benda, C., Dunzinger, S., Huang, Y., Ho, J. C., Yang, J., Wang, Y., Zhang, Y., Zhuang, Q., Li, Y., Bao, X., Tse, H.-F., Grillari, J., Grillari-Voglauer, R., Pei, D. and Esteban, M. A. (2012) 'Generation of human induced pluripotent stem cells from urine samples.' *Nature Protocols*, 7(12) pp. 2080–2089.